

B4

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
18 April 2002 (18.04.2002)

PCT

(10) International Publication Number
WO 02/30950 A2

- (51) International Patent Classification⁷: **C07K**
- (21) International Application Number: **PCT/US01/42725**
- (22) International Filing Date: 15 October 2001 (15.10.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/240,275 13 October 2000 (13.10.2000) US
- (71) Applicant (for all designated States except US): **GENAIS-SANCE PHARMACEUTICALS, INC.** [US/US]; A Corporation of the State of Delaware, Five Science Park, New Haven, CT 06511 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **CHEW, Anne** [US/US]; 1477 Beacon Street, #64, Brookline, MA 02446 (US). **CHOI, Julie, Y.** [US/US]; 38 Elizabeth Street, West Haven, CT 06516 (US). **KOSHY, Beena** [IN/US]; 1298 Hartford Turnpike, Apt. 11B, North Haven, CT 06473 (US).
- (74) Agents: **FIELD, Gisela, M. et al.**; Genaissance Pharmaceuticals, Inc., Five Science Park, New Haven, CT 06511 (US).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:
— without international search report and to be republished upon receipt of that report
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 02/30950 A2

(54) Title: HAPLOTYPES OF THE FY GENE

(57) Abstract: Novel genetic variants of the Duffy Blood Group (FY) gene are described. Various genotypes, haplotypes, and haplotype pairs that exist in the general United States population are disclosed for the FY gene. Compositions and methods for haplotyping and/or genotyping the FY gene in an individual are also disclosed. Polynucleotides defined by the haplotypes disclosed herein are also described.

HAPLOTYPES OF THE FY GENE

RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application Serial No. 60/240,275 filed
5 October 13, 2000.

FIELD OF THE INVENTION

This invention relates to variation in genes that encode pharmaceutically-important proteins.
In particular, this invention provides genetic variants of the human Duffy blood group (FY) gene and
10 methods for identifying which variant(s) of this gene is/are possessed by an individual.

BACKGROUND OF THE INVENTION

Current methods for identifying pharmaceuticals to treat disease often start by identifying,
cloning, and expressing an important target protein related to the disease. A determination of whether
15 an agonist or antagonist is needed to produce an effect that may benefit a patient with the disease is
then made. Then, vast numbers of compounds are screened against the target protein to find new
potential drugs. The desired outcome of this process is a lead compound that is specific for the target,
thereby reducing the incidence of the undesired side effects usually caused by activity at non-intended
targets. The lead compound identified in this screening process then undergoes further *in vitro* and *in*
20 *vivo* testing to determine its absorption, disposition, metabolism and toxicological profiles. Typically,
this testing involves use of cell lines and animal models with limited, if any, genetic diversity.

What this approach fails to consider, however, is that natural genetic variability exists between
individuals in any and every population with respect to pharmaceutically-important proteins, including
the protein targets of candidate drugs, the enzymes that metabolize these drugs and the proteins whose
25 activity is modulated by such drug targets. Subtle alteration(s) in the primary nucleotide sequence of a
gene encoding a pharmaceutically-important protein may be manifested as significant variation in
expression, structure and/or function of the protein. Such alterations may explain the relatively high
degree of uncertainty inherent in the treatment of individuals with a drug whose design is based upon a
single representative example of the target or enzyme(s) involved in metabolizing the drug. For
30 example, it is well-established that some drugs frequently have lower efficacy in some individuals than
others, which means such individuals and their physicians must weigh the possible benefit of a larger
dosage against a greater risk of side effects. Also, there is significant variation in how well people
metabolize drugs and other exogenous chemicals, resulting in substantial interindividual variation in
the toxicity and/or efficacy of such exogenous substances (Evans et al., 1999, *Science* 286:487-491).
35 This variability in efficacy or toxicity of a drug in genetically-diverse patients makes many drugs
ineffective or even dangerous in certain groups of the population, leading to the failure of such drugs in
clinical trials or their early withdrawal from the market even though they could be highly beneficial for

other groups in the population. This problem significantly increases the time and cost of drug discovery and development, which is a matter of great public concern.

It is well-recognized by pharmaceutical scientists that considering the impact of the genetic variability of pharmaceutically-important proteins in the early phases of drug discovery and development is likely to reduce the failure rate of candidate and approved drugs (Marshall A 1997 *Nature Biotech* 15:1249-52; Kleyn PW et al. 1998 *Science* 281: 1820-21; Kola I 1999 *Curr Opin Biotech* 10:589-92; Hill AVS et al. 1999 in *Evolution in Health and Disease* Stearns SS (Ed.) Oxford University Press, New York, pp 62-76; Meyer U.A. 1999 in *Evolution in Health and Disease* Stearns SS (Ed.) Oxford University Press, New York, pp 41-49; Kalow W et al. 1999 *Clin. Pharm. Therap.* 66:445-7; Marshall, E 1999 *Science* 284:406-7; Judson R et al. 2000 *Pharmacogenomics* 1:1-12; Roses AD 2000 *Nature* 405:857-65). However, in practice this has been difficult to do, in large part because of the time and cost required for discovering the amount of genetic variation that exists in the population (Chakravarti A 1998 *Nature Genet* 19:216-7; Wang DG et al 1998 *Science* 280:1077-82; Chakravarti A 1999 *Nat Genet* 21:56-60 (suppl); Stephens JC 1999 *Mol. Diagnosis* 4:309-317; Kwok PY and Gu S 1999 *Mol. Med. Today* 5:538-43; Davidson S 2000 *Nature Biotech* 18:1134-5).

The standard for measuring genetic variation among individuals is the haplotype, which is the ordered combination of polymorphisms in the sequence of each form of a gene that exists in the population. Because haplotypes represent the variation across each form of a gene, they provide a more accurate and reliable measurement of genetic variation than individual polymorphisms. For example, while specific variations in gene sequences have been associated with a particular phenotype such as disease susceptibility (Roses AD *supra*; Ulbrecht M et al. 2000 *Am J Respir Crit Care Med* 161: 469-74) and drug response (Wolfe CR et al. 2000 *BMJ* 320:987-90; Dahl BS 1997 *Acta Psychiatr Scand* 96 (Suppl 391): 14-21), in many other cases an individual polymorphism may be found in a variety of genomic backgrounds, i.e., different haplotypes, and therefore shows no definitive coupling between the polymorphism and the causative site for the phenotype (Clark AG et al. 1998 *Am J Hum Genet* 63:595-612; Ulbrecht M et al. 2000 *supra*; Drysdale et al. 2000 *PNAS* 97:10483-10488). Thus, there is an unmet need in the pharmaceutical industry for information on what haplotypes exist in the population for pharmaceutically-important genes. Such haplotype information would be useful in improving the efficiency and output of several steps in the drug discovery and development process, including target validation, identifying lead compounds, and early phase clinical trials (Marshall et al., *supra*).

One pharmaceutically-important gene for the treatment of malaria and inflammatory diseases is the Duffy blood group (FY) gene or its encoded product. FY, also known as DARC, is a Duffy blood group associated glycoprotein that carries Duffy blood group antigens (Iwamoto et al. *Blood* 1995 Feb 1;85(3):622-6). The Duffy blood group antigens have been characterized by their roles as receptors on red blood cells for the malarial parasites and as promiscuous receptors for the chemokine superfamily. Malaria parasites, such as *Plasmodium vivax*, only enter red blood cells when the FY

protein is present (Rios and Bianco, *Semin Hematol* 2000; 37(2):177-85). The parasite-specific binding site, the binding site of chemokines, and the major antigenic domains of the FY protein are located in overlapping regions at the extracellular N-terminus of the FY protein (Pogo and Chaudhuri, *Semin Hematol* 2000; 37(2):122-9). Thus, FY has been associated with malaria and may be involved in regulation of the level of circulating proinflammatory chemokines (Woolley et al., *Transfusion* 2000; 40(8):949-53).

The Duffy blood group gene is located on chromosome 1q21-q25 and encodes two alternately spliced proteins. One spliced form of the mRNA yields a 338 amino acid protein, which was the first described form of FY (Figure 3). It was later discovered that alternative splicing of the mRNA gives rise to a 336 amino acid form of the protein (not shown)(Iwamoto et al., *Blood* 1996; 87:378-85). A reference sequence for the FY gene is shown in the contiguous lines of Figure 1 (Genaissance Reference No. 4118011; SEQ ID NO: 1). Reference sequences for the coding sequence (GenBank Accession No. NM_002036.1) and protein are shown in Figures 2 (SEQ ID NO: 2) and 3 (SEQ ID NO: 3), respectively.

Several polymorphisms in the FY gene have been reported. Tournamille et al. (*Nat Genet* 1995; 10(2):224-8) discovered that the molecular basis of the Fy (a-b-) phenotype was due to a polymorphism of thymine or cytosine in the promoter region of the FY gene. This polymorphism corresponds to nucleotide 3470 in Figure 1, and is herein referred to as PS9. The presence of this polymorphism, which is common in people of African origin but rare in other ethnic groups, results in the absence of the FY glycoprotein from red cells and, therefore, resistance to *P. vivax* or malarial infection (Tournamille et al., *Nat Genet* 1995; 10(2):224-8; Daniels, *Transfus Clin Biol* 1997; 4(4):383-90). Another polymorphism in the FY gene is that of an adenine or guanine at a position corresponding to nucleotide position 4140 in Figure 1, herein referred to as PS14. This polymorphism results in an aspartic acid or glycine amino acid variation (Tournamille et al., *Hum Genet* 1995; 95(4):407-410). There are two other polymorphisms in the FY gene that cause amino acid variations in the FY protein. These polymorphisms include a cytosine or thymine at a position corresponding to nucleotide position 4280 (herein referred to as PS16) and a guanine or adenine at a position corresponding to nucleotide position 4313 (herein referred to as PS17) in Figure 1 (Yazdanbakhsh et al. *Transfusion* 2000 Mar;40(3):310-20). These polymorphisms result in an arginine or cysteine amino acid variation at a position corresponding to amino acid position 91 (R91C) and an alanine or threonine amino acid variation at a position corresponding to amino acid position 102 (A102C) in Figure 3, respectively. In the case of the R91C amino acid variation, the polymorphism results in a considerable change to the chemical nature of the protein, suggesting that this polymorphism may affect antigenic determinants of FY, and therefore, may be of clinical significance (Parasol et al., *Blood* 92: 2237-2243).

Because of the potential for variation in the FY gene to affect the expression and function of the encoded protein, it would be useful to know whether additional polymorphisms exist in the FY

gene, as well as how such polymorphisms are combined in different copies of the gene. Such information could be applied for studying the biological function of FY as well as in identifying drugs targeting this protein for the treatment of disorders related to its abnormal expression or function.

5 SUMMARY OF THE INVENTION

Accordingly, the inventors herein have discovered 16 novel polymorphic sites in the FY gene. These polymorphic sites (PS) correspond to the following nucleotide positions in Figure 1: 2690 (PS1), 2864 (PS2), 2882 (PS3), 2910 (PS4), 2949 (PS5), 2980 (PS6), 2996 (PS7), 3259 (PS8), 3672 (PS10), 3707 (PS11), 3979 (PS12), 3997 (PS13), 4214 (PS15), 4617 (PS18), 4618 (PS19) and 4992 (PS20). The polymorphisms at these sites are cytosine or thymine at PS1, guanine or adenine at PS2, adenine or guanine at PS3, cytosine or thymine at PS4, cytosine or adenine at PS5, guanine or cytosine at PS6, cytosine or thymine at PS7, thymine or cytosine at PS8, cytosine or thymine at PS10, cytosine or thymine at PS11, adenine or guanine at PS12, cytosine or thymine at PS13, cytosine or thymine at PS15, cytosine or thymine at PS18, guanine or adenine at PS19 and cytosine or thymine at PS20. In addition, the inventors have determined the identity of the alleles at these sites, as well as at the previously identified sites at nucleotide positions 3470 (PS9), 4140 (PS14), 4280 (PS16) and 4313 (PS17), in a human reference population of 79 unrelated individuals self-identified as belonging to one of four major population groups: African descent, Asian, Caucasian and Hispanic/Latino. From this information, the inventors deduced a set of haplotypes and haplotype pairs for PS1-PS20 in the FY gene, which are shown below in Tables 4 and 3, respectively. Each of these FY haplotypes constitutes a code that defines the variant nucleotides that exist in the human population at this set of polymorphic sites in the FY gene. Thus each FY haplotype also represents a naturally-occurring isoform (also referred to herein as an "isogene") of the FY gene. The frequency of each haplotype and haplotype pair within the total reference population and within each of the four major population groups included in the reference population was also determined.

Thus, in one embodiment, the invention provides a method, composition and kit for genotyping the FY gene in an individual. The genotyping method comprises identifying the nucleotide pair that is present at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS10, PS11, PS12, PS13, PS15, PS18, PS19 and PS20 in both copies of the FY gene from the individual. A genotyping composition of the invention comprises an oligonucleotide probe or primer which is designed to specifically hybridize to a target region containing, or adjacent to, one of these novel FY polymorphic sites. A genotyping kit of the invention comprises a set of oligonucleotides designed to genotype each of these novel FY polymorphic sites. In a preferred embodiment, the genotyping kit comprises a set of oligonucleotides designed to genotype each of PS1-PS20. The genotyping method, composition, and kit are useful in determining whether an individual has one of the haplotypes in Table 4 below or has one of the haplotype pairs in Table 3 below.

The invention also provides a method for haplotyping the FY gene in an individual. In one embodiment, the haplotyping method comprises determining, for one copy of the FY gene, the identity of the nucleotide at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS10, PS11, PS12, PS13, PS15, PS18, PS19 and PS20. In another embodiment, the haplotyping method comprises determining whether one copy of the individual's FY gene is defined by one of the FY haplotypes shown in Table 4, below, or a sub-haplotype thereof. In a preferred embodiment, the haplotyping method comprises determining whether both copies of the individual's FY gene are defined by one of the FY haplotype pairs shown in Table 3 below, or a sub-haplotype pair thereof. Establishing the FY haplotype or haplotype pair of an individual is useful for improving the efficiency and reliability of several steps in the discovery and development of drugs for treating diseases associated with FY activity, e.g., malaria and inflammatory disorders

For example, the haplotyping method can be used by the pharmaceutical research scientist to validate FY as a candidate target for treating a specific condition or disease predicted to be associated with FY activity. Determining for a particular population the frequency of one or more of the individual FY haplotypes or haplotype pairs described herein will facilitate a decision on whether to pursue FY as a target for treating the specific disease of interest. In particular, if variable FY activity is associated with the disease, then one or more FY haplotypes or haplotype pairs will be found at a higher frequency in disease cohorts than in appropriately genetically matched controls. Conversely, if each of the observed FY haplotypes are of similar frequencies in the disease and control groups, then it may be inferred that variable FY activity has little, if any, involvement with that disease. In either case, the pharmaceutical research scientist can, without *a priori* knowledge as to the phenotypic effect of any FY haplotype or haplotype pair, apply the information derived from detecting FY haplotypes in an individual to decide whether modulating FY activity would be useful in treating the disease.

The claimed invention is also useful in screening for compounds targeting FY to treat a specific condition or disease predicted to be associated with FY activity. For example, detecting which of the FY haplotypes or haplotype pairs disclosed herein are present in individual members of a population with the specific disease of interest enables the pharmaceutical scientist to screen for a compound(s) that displays the highest desired agonist or antagonist activity for each of the FY isoforms present in the disease population, or for only the most frequent FY isoforms present in the disease population. Thus, without requiring any *a priori* knowledge of the phenotypic effect of any particular FY haplotype or haplotype pair, the claimed haplotyping method provides the scientist with a tool to identify lead compounds that are more likely to show efficacy in clinical trials.

Haplotyping the FY gene in an individual is also useful in the design of clinical trials of candidate drugs for treating a specific condition or disease predicted to be associated with FY activity. For example, instead of randomly assigning patients with the disease of interest to the treatment or control group as is typically done now, determining which of the FY haplotype(s) disclosed herein are present in individual patients enables the pharmaceutical scientist to distribute FY haplotypes and/or

haplotype pairs evenly to treatment and control groups, thereby reducing the potential for bias in the results that could be introduced by a larger frequency of a FY haplotype or haplotype pair that is associated with response to the drug being studied in the trial, even if this association was previously unknown. Thus, by practicing the claimed invention, the scientist can more confidently rely on the information learned from the trial, without first determining the phenotypic effect of any FY haplotype or haplotype pair.

In another embodiment, the invention provides a method for identifying an association between a trait and a FY genotype, haplotype, or haplotype pair for one or more of the novel polymorphic sites described herein. The method comprises comparing the frequency of the FY genotype, haplotype, or haplotype pair in a population exhibiting the trait with the frequency of the FY genotype or haplotype in a reference population. A higher frequency of the FY genotype, haplotype, or haplotype pair in the trait population than in the reference population indicates the trait is associated with the FY genotype, haplotype, or haplotype pair. In preferred embodiments, the trait is susceptibility to a disease, severity of a disease, the staging of a disease or response to a drug. In a particularly preferred embodiment, the FY haplotype is selected from the haplotypes shown in Table 4, or a sub-haplotype thereof. Such methods have applicability in developing diagnostic tests and therapeutic treatments for malaria and inflammatory disorders.

In yet another embodiment, the invention provides an isolated polynucleotide comprising a nucleotide sequence which is a polymorphic variant of a reference sequence for the FY gene or a fragment thereof. The reference sequence comprises the contiguous sequences shown in Figure 1 and the polymorphic variant comprises at least one polymorphism selected from the group consisting of thymine at PS1, adenine at PS2, guanine at PS3, thymine at PS4, adenine at PS5, cytosine at PS6, thymine at PS7, cytosine at PS8, thymine at PS10, thymine at PS11, guanine at PS12, thymine at PS13, thymine at PS15, thymine at PS18, adenine at PS19 and thymine at PS20. In a preferred embodiment, the polymorphic variant comprises one or more additional polymorphisms selected from the group consisting of cytosine at PS9, guanine at PS14, thymine at PS16 and adenine at PS17.

A particularly preferred polymorphic variant is an isogene of the FY gene. A FY isogene of the invention comprises cytosine or thymine at PS1, guanine or adenine at PS2, adenine or guanine at PS3, cytosine or thymine at PS4, cytosine or adenine at PS5, guanine or cytosine at PS6, cytosine or thymine at PS7, thymine or cytosine at PS8, thymine or cytosine at PS9, cytosine or thymine at PS10, cytosine or thymine at PS11, adenine or guanine at PS12, cytosine or thymine at PS13, adenine or guanine at PS14, cytosine or thymine at PS15, cytosine or thymine at PS16, guanine or adenine at PS17, cytosine or thymine at PS18, guanine or adenine at PS19 and cytosine or thymine at PS20. The invention also provides a collection of FY isogenes, referred to herein as a FY genome anthology.

In another embodiment, the invention provides a polynucleotide comprising a polymorphic variant of a reference sequence for a FY cDNA or a fragment thereof. The reference sequence comprises SEQ ID NO:2 (Fig.2) and the polymorphic cDNA comprises at least one polymorphism

selected from the group consisting of thymine at a position corresponding to nucleotide 205, thymine at a position corresponding to nucleotide 608, adenine at a position corresponding to nucleotide 609 and thymine at a position corresponding to nucleotide 983. In a preferred embodiment, the polymorphic variant comprises one or more additional polymorphisms selected from the group consisting of guanine at a position corresponding to nucleotide 131, thymine at a position corresponding to nucleotide 271 and adenine at a position corresponding to nucleotide 304. A particularly preferred polymorphic cDNA variant comprises the coding sequence of a FY isogene defined by haplotypes 2-4, 6, 9, 10-12, and 16-18.

Polynucleotides complementary to these FY genomic and cDNA variants are also provided by the invention. It is believed that polymorphic variants of the FY gene will be useful in studying the expression and function of FY, and in expressing FY protein for use in screening for candidate drugs to treat diseases related to FY activity.

In other embodiments, the invention provides a recombinant expression vector comprising one of the polymorphic genomic and cDNA variants operably linked to expression regulatory elements as well as a recombinant host cell transformed or transfected with the expression vector. The recombinant vector and host cell may be used to express FY for protein structure analysis and drug binding studies.

In yet another embodiment, the invention provides a polypeptide comprising a polymorphic variant of a reference amino acid sequence for the FY protein. The reference amino acid sequence comprises SEQ ID NO:3 (Fig.3) and the polymorphic variant comprises at least one variant amino acid selected from the group consisting of phenylalanine at a position corresponding to amino acid position 69, isoleucine at a position corresponding to amino acid position 203, isoleucine at a position corresponding to amino acid position 203 and phenylalanine at a position corresponding to amino acid position 328. In some embodiments, the polymorphic variant also comprises at least one variant amino acid selected from the group consisting of glycine at a position corresponding to amino acid position 44, cysteine at a position corresponding to amino acid position 91 and threonine at a position corresponding to amino acid position 102. A polymorphic variant of FY is useful in studying the effect of the variation on the biological activity of FY as well as on the binding affinity of candidate drugs targeting FY for the treatment of malaria and inflammatory disorders.

The present invention also provides antibodies that recognize and bind to the above polymorphic FY protein variant. Such antibodies can be utilized in a variety of diagnostic and prognostic formats and therapeutic methods.

The present invention also provides nonhuman transgenic animals comprising one or more of the FY polymorphic genomic variants described herein and methods for producing such animals. The transgenic animals are useful for studying expression of the FY isogenes *in vivo*, for *in vivo* screening and testing of drugs targeted against FY protein, and for testing the efficacy of therapeutic agents and compounds for in a biological system.

The present invention also provides a computer system for storing and displaying polymorphism data determined for the FY gene. The computer system comprises a computer processing unit; a display; and a database containing the polymorphism data. The polymorphism data includes one or more of the following: the polymorphisms, the genotypes, the haplotypes, and the haplotype pairs identified for the FY gene in a reference population. In a preferred embodiment, the computer system is capable of producing a display showing FY haplotypes organized according to their evolutionary relationships.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates a reference sequence for the FY gene (Genaissance Reference No. 4118011; contiguous lines), with the start and stop positions of each region of coding sequence indicated with a bracket ([or]) and the numerical position below the sequence and the polymorphic site(s) and polymorphism(s) identified by Applicants in a reference population indicated by the variant nucleotide positioned below the polymorphic site in the sequence. SEQ ID NO:1 is equivalent to Figure 1, with the two alternative allelic variants of each polymorphic site indicated by the appropriate nucleotide symbol (R= G or A, Y= T or C, M= A or C, K= G or T, S= G or C, and W= A or T; WIPO standard ST.25). SEQ ID NO:84 is a modified version of SEQ ID NO:1 that shows the context sequence of each polymorphic site, PS1-PS20, in a uniform format to facilitate electronic searching. For each polymorphic site, SEQ ID NO:84 contains a block of 60 bases of the nucleotide sequence encompassing the centrally-located polymorphic site at the 30th position, followed by 60 bases of unspecified sequence to represent that each PS is separated by genomic sequence whose composition is defined elsewhere herein.

Figure 2 illustrates a reference sequence for the FY coding sequence (contiguous lines; SEQ ID NO:2), with the polymorphic site(s) and polymorphism(s) identified by Applicants in a reference population indicated by the variant nucleotide positioned below the polymorphic site in the sequence.

Figure 3 illustrates a reference sequence for the FY protein (contiguous lines; SEQ ID NO:3), with the variant amino acid(s) caused by the polymorphism(s) of Figure 2 positioned below the polymorphic site in the sequence.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is based on the discovery of novel variants of the FY gene. As described in more detail below, the inventors herein discovered 23 isogenes of the FY gene by characterizing the FY gene found in genomic DNAs isolated from an Index Repository that contains immortalized cell lines from one chimpanzee and 93 human individuals. The human individuals included a reference population of 79 unrelated individuals self-identified as belonging to one of four major population groups: Caucasian (21 individuals), African descent (20 individuals), Asian (20 individuals), or Hispanic/Latino (18 individuals). To the extent possible, the members of this reference

population were organized into population subgroups by their self-identified ethnogeographic origin as shown in Table 1 below. In addition, the Index Repository contains three unrelated indigenous American Indians (one from each of North, Central and South America), one three-generation Caucasian family (from the CEPH Utah cohort) and one two-generation African-American family.

5

Table 1. Population Groups in the Index Repository

Population Group	Population Subgroup	No. of Individuals
African descent		20
	Sierra Leone	1
Asian		20
	Burma	1
	China	3
	Japan	6
	Korea	1
	Philippines	5
	Vietnam	4
Caucasian		21
	British Isles	3
	British Isles/Central	4
	British Isles/Eastern	1
	Central/Eastern	1
	Eastern	3
	Central/Mediterranean	1
	Mediterranean	2
	Scandinavian	2
Hispanic/Latino		18
	Caribbean	8
	Caribbean (Spanish Descent)	2
	Central American (Spanish Descent)	1
	Mexican American	4
	South American (Spanish Descent)	3

The FY isogenes present in the human reference population are defined by haplotypes for 20 polymorphic sites in the FY gene, 16 of which are believed to be novel. The FY polymorphic sites identified by the inventors are referred to as PS1-PS20 to designate the order in which they are located in the gene (see Table 2 below), with the novel polymorphic sites referred to as PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS10, PS11, PS12, PS13, PS15, PS18, PS19 and PS20. Using the genotypes identified in the Index Repository for PS1-PS20 and the methodology described in the Examples below, the inventors herein also determined the pair of haplotypes for the FY gene present in individual human members of this repository. The human genotypes and haplotypes found in the repository for the FY gene include those shown in Tables 3 and 4, respectively. The polymorphism and haplotype data disclosed herein are useful for validating whether FY is a suitable target for drugs to treat, screening for such drugs and reducing bias in clinical trials of such drugs.

In the context of this disclosure, the following terms shall be defined as follows unless otherwise indicated:

Allele - A particular form of a genetic locus, distinguished from other forms by its particular nucleotide sequence.

Candidate Gene - A gene which is hypothesized to be responsible for a disease, condition, or the response to a treatment, or to be correlated with one of these.

5 **Gene** - A segment of DNA that contains all the information for the regulated biosynthesis of an RNA product, including promoters, exons, introns, and other untranslated regions that control expression.

10 **Genotype** - An unphased 5' to 3' sequence of nucleotide pair(s) found at one or more polymorphic sites in a locus on a pair of homologous chromosomes in an individual. As used herein, genotype includes a full-genotype and/or a sub-genotype as described below.

Full-genotype - The unphased 5' to 3' sequence of nucleotide pairs found at all polymorphic sites examined herein in a locus on a pair of homologous chromosomes in a single individual.

15 **Sub-genotype** - The unphased 5' to 3' sequence of nucleotides seen at a subset of the polymorphic sites examined herein in a locus on a pair of homologous chromosomes in a single individual.

Genotyping - A process for determining a genotype of an individual.

Haplotype - A 5' to 3' sequence of nucleotides found at one or more polymorphic sites in a locus on a single chromosome from a single individual. As used herein, haplotype includes a full-haplotype and/or a sub-haplotype as described below.

20 **Full-haplotype** - The 5' to 3' sequence of nucleotides found at all polymorphic sites examined herein in a locus on a single chromosome from a single individual.

Sub-haplotype - The 5' to 3' sequence of nucleotides seen at a subset of the polymorphic sites examined herein in a locus on a single chromosome from a single individual.

Haplotype pair - The two haplotypes found for a locus in a single individual.

25 **Haplotyping** - A process for determining one or more haplotypes in an individual and includes use of family pedigrees, molecular techniques and/or statistical inference.

30 **Haplotype data** - Information concerning one or more of the following for a specific gene: a listing of the haplotype pairs in each individual in a population; a listing of the different haplotypes in a population; frequency of each haplotype in that or other populations, and any known associations between one or more haplotypes and a trait.

Isoform - A particular form of a gene, mRNA, cDNA, coding sequence or the protein-encoded thereby, distinguished from other forms by its particular sequence and/or structure.

Isogene - One of the isoforms (e.g., alleles) of a gene found in a population. An isogene (or allele) contains all of the polymorphisms present in the particular isoform of the gene.

35 **Isolated** - As applied to a biological molecule such as RNA, DNA, oligonucleotide, or protein, isolated means the molecule is substantially free of other biological molecules such as nucleic acids, proteins, lipids, carbohydrates, or other material such as cellular debris and growth media. Generally,

the term "isolated" is not intended to refer to a complete absence of such material or to absence of water, buffers, or salts, unless they are present in amounts that substantially interfere with the methods of the present invention.

5 **Locus** - A location on a chromosome or DNA molecule corresponding to a gene or a physical or phenotypic feature, where physical features include polymorphic sites.

Naturally-occurring - A term used to designate that the object it is applied to, e.g., naturally-occurring polynucleotide or polypeptide, can be isolated from a source in nature and which has not been intentionally modified by man.

10 **Nucleotide pair** - The nucleotides found at a polymorphic site on the two copies of a chromosome from an individual.

Phased - As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, phased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is known.

15 **Polymorphic site (PS)** - A position on a chromosome or DNA molecule at which at least two alternative sequences are found in a population.

Polymorphic variant (variant) - A gene, mRNA, cDNA, polypeptide, protein or peptide whose nucleotide or amino acid sequence varies from a reference sequence due to the presence of a polymorphism in the gene.

20 **Polymorphism** - The sequence variation observed in an individual at a polymorphic site. Polymorphisms include nucleotide substitutions, insertions, deletions and microsatellites and may, but need not, result in detectable differences in gene expression or protein function.

Polymorphism data - Information concerning one or more of the following for a specific gene: location of polymorphic sites; sequence variation at those sites; frequency of polymorphisms in one or more populations; the different genotypes and/or haplotypes determined for the gene; frequency
25 of one or more of these genotypes and/or haplotypes in one or more populations; any known association(s) between a trait and a genotype or a haplotype for the gene.

Polymorphism Database - A collection of polymorphism data arranged in a systematic or methodical way and capable of being individually accessed by electronic or other means.

30 **Polynucleotide** - A nucleic acid molecule comprised of single-stranded RNA or DNA or comprised of complementary, double-stranded DNA.

Population Group - A group of individuals sharing a common ethnogeographic origin.

Reference Population - A group of subjects or individuals who are predicted to be representative of the genetic variation found in the general population. Typically, the reference population represents the genetic variation in the population at a certainty level of at least 85%,
35 preferably at least 90%, more preferably at least 95% and even more preferably at least 99%.

Single Nucleotide Polymorphism (SNP) - Typically, the specific pair of nucleotides observed at a single polymorphic site. In rare cases, three or four nucleotides may be found.

Subject – A human individual whose genotypes or haplotypes or response to treatment or disease state are to be determined.

Treatment – A stimulus administered internally or externally to a subject.

Unphased – As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, unphased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is not known.

As discussed above, information on the identity of genotypes and haplotypes for the FY gene of any particular individual as well as the frequency of such genotypes and haplotypes in any particular population of individuals is useful for a variety of drug discovery and development applications. Thus, the invention also provides compositions and methods for detecting the novel FY polymorphisms, haplotypes and haplotype pairs identified herein.

The compositions comprise at least one oligonucleotide for detecting the variant nucleotide or nucleotide pair located at a novel FY polymorphic site in one copy or two copies of the FY gene. Such oligonucleotides are referred to herein as FY haplotyping oligonucleotides or genotyping oligonucleotides, respectively, and collectively as FY oligonucleotides. In one embodiment, a FY haplotyping or genotyping oligonucleotide is a probe or primer capable of hybridizing to a target region that contains, or that is located close to, one of the novel polymorphic sites described herein.

As used herein, the term “oligonucleotide” refers to a polynucleotide molecule having less than about 100 nucleotides. A preferred oligonucleotide of the invention is 10 to 35 nucleotides long. More preferably, the oligonucleotide is between 15 and 30, and most preferably, between 20 and 25 nucleotides in length. The exact length of the oligonucleotide will depend on many factors that are routinely considered and practiced by the skilled artisan. The oligonucleotide may be comprised of any phosphorylation state of ribonucleotides, deoxyribonucleotides, and acyclic nucleotide derivatives, and other functionally equivalent derivatives. Alternatively, oligonucleotides may have a phosphate-free backbone, which may be comprised of linkages such as carboxymethyl, acetamidate, carbamate, polyamide (peptide nucleic acid (PNA)) and the like (Varma, R. in Molecular Biology and Biotechnology, A Comprehensive Desk Reference, Ed. R. Meyers, VCH Publishers, Inc. (1995), pages 617-620). Oligonucleotides of the invention may be prepared by chemical synthesis using any suitable methodology known in the art, or may be derived from a biological sample, for example, by restriction digestion. The oligonucleotides may be labeled, according to any technique known in the art, including use of radiolabels, fluorescent labels, enzymatic labels, proteins, haptens, antibodies, sequence tags and the like.

Haplotyping or genotyping oligonucleotides of the invention must be capable of specifically hybridizing to a target region of a FY polynucleotide. Preferably, the target region is located in a FY isogene. As used herein, specific hybridization means the oligonucleotide forms an anti-parallel double-stranded structure with the target region under certain hybridizing conditions, while failing to form such a structure when incubated with another region in the FY polynucleotide or with a non-FY

polynucleotide under the same hybridizing conditions. Preferably, the oligonucleotide specifically hybridizes to the target region under conventional high stringency conditions. The skilled artisan can readily design and test oligonucleotide probes and primers suitable for detecting polymorphisms in the FY gene using the polymorphism information provided herein in conjunction with the known sequence information for the FY gene and routine techniques.

A nucleic acid molecule such as an oligonucleotide or polynucleotide is said to be a "perfect" or "complete" complement of another nucleic acid molecule if every nucleotide of one of the molecules is complementary to the nucleotide at the corresponding position of the other molecule. A nucleic acid molecule is "substantially complementary" to another molecule if it hybridizes to that molecule with sufficient stability to remain in a duplex form under conventional low-stringency conditions. Conventional hybridization conditions are described, for example, by Sambrook J. et al., in *Molecular Cloning, A Laboratory Manual*, 2nd Edition, Cold Spring Harbor Press, Cold Spring Harbor, NY (1989) and by Haymes, B.D. et al. in *Nucleic Acid Hybridization, A Practical Approach*, IRL Press, Washington, D.C. (1985). While perfectly complementary oligonucleotides are preferred for detecting polymorphisms, departures from complete complementarity are contemplated where such departures do not prevent the molecule from specifically hybridizing to the target region. For example, an oligonucleotide primer may have a non-complementary fragment at its 5' end, with the remainder of the primer being complementary to the target region. Alternatively, non-complementary nucleotides may be interspersed into the probe or primer as long as the resulting probe or primer is still capable of specifically hybridizing to the target region.

Preferred haplotyping or genotyping oligonucleotides of the invention are allele-specific oligonucleotides. As used herein, the term allele-specific oligonucleotide (ASO) means an oligonucleotide that is able, under sufficiently stringent conditions, to hybridize specifically to one allele of a gene, or other locus, at a target region containing a polymorphic site while not hybridizing to the corresponding region in another allele(s). As understood by the skilled artisan, allele-specificity will depend upon a variety of readily optimized stringency conditions, including salt and formamide concentrations, as well as temperatures for both the hybridization and washing steps. Examples of hybridization and washing conditions typically used for ASO probes are found in Kogan et al., "Genetic Prediction of Hemophilia A" in *PCR Protocols, A Guide to Methods and Applications*, Academic Press, 1990 and Ruaño et al., 87 *Proc. Natl. Acad. Sci. USA* 6296-6300, 1990. Typically, an ASO will be perfectly complementary to one allele while containing a single mismatch for another allele.

Allele-specific oligonucleotides of the invention include ASO probes and ASO primers. ASO probes which usually provide good discrimination between different alleles are those in which a central position of the oligonucleotide probe aligns with the polymorphic site in the target region (e.g., approximately the 7th or 8th position in a 15mer, the 8th or 9th position in a 16mer, and the 10th or 11th position in a 20mer). An ASO primer of the invention has a 3' terminal nucleotide, or preferably a 3'

penultimate nucleotide, that is complementary to only one nucleotide of a particular SNP, thereby acting as a primer for polymerase-mediated extension only if the allele containing that nucleotide is present. ASO probes and primers hybridizing to either the coding or noncoding strand are contemplated by the invention. ASO probes and primers listed below use the appropriate nucleotide symbol (R= G or A, Y= T or C, M= A or C, K= G or T, S= G or C, and W= A or T; WIPO standard ST.25) at the position of the polymorphic site to represent that the ASO contains either of the two alternative allelic variants observed at that polymorphic site.

A preferred ASO probe for detecting FY gene polymorphisms comprises a nucleotide sequence, listed 5' to 3', selected from the group consisting of:

10 TGTCAGAYCATGTAT (SEQ ID NO:4) and its complement,
 GACACCCRCCAAGCC (SEQ ID NO:5) and its complement,
 CACATACRGATATGT (SEQ ID NO:6) and its complement,
 CAGCAAAYGTACACA (SEQ ID NO:7) and its complement,
 15 ACGCCCAMGTGCACA (SEQ ID NO:8) and its complement,
 CAGAGTTSACCACCA (SEQ ID NO:9) and its complement,
 CACCTTTYTCCCAA (SEQ ID NO:10) and its complement,
 TCTCCCTYTCCACTT (SEQ ID NO:11) and its complement,
 CCCTTCCYGTCTTTT (SEQ ID NO:12) and its complement,
 20 TTTCTCTYTCTCCTT (SEQ ID NO:13) and its complement,
 CCCCTCARTTCCCAG (SEQ ID NO:14) and its complement,
 ACTCTTCYGGTGTA (SEQ ID NO:15) and its complement,
 CTTTCATCYTCACCAG (SEQ ID NO:16) and its complement,
 TACAGCAYGGAGCTG (SEQ ID NO:17) and its complement,
 25 ACAGCACRGAGCTGA (SEQ ID NO:18) and its complement, and
 GGATGGTYTTCTCAT (SEQ ID NO:19) and its complement.

A preferred ASO primer for detecting FY gene polymorphisms comprises a nucleotide sequence, listed 5' to 3', selected from the group consisting of:

30 TTAACCTTGTCAGAYC (SEQ ID NO:20); AGTGAATACATGRT (SEQ ID NO:21);
 CACCCAGACACCCRC (SEQ ID NO:22); GTGAGGGGCTTGGYG (SEQ ID NO:23);
 GCCCCTCACATACRG (SEQ ID NO:24); TTGTGCACATATCYG (SEQ ID NO:25);
 GATACACAGCAAAYG (SEQ ID NO:26); GAACTCTGTGTACRT (SEQ ID NO:27);
 35 GAGCTCAGCCCCAMG (SEQ ID NO:28); GGGGTGTGTGCACT (SEQ ID NO:29);
 TTGGGACAGAGTTSA (SEQ ID NO:30); AGGTGGTGGTGGTSA (SEQ ID NO:31);
 CACCACCACCTTTYT (SEQ ID NO:32); CATGTGTTTGGGARA (SEQ ID NO:33);
 TCTCAATCTCCCTYT (SEQ ID NO:34); TTACCGAAGTGGARA (SEQ ID NO:35);
 TTTTCTCCCTTCCYG (SEQ ID NO:36); AGAGGAAAAAGCRG (SEQ ID NO:37);
 40 AGTCTTTTTCCTTYT (SEQ ID NO:38); CATAGGAAGGAGARA (SEQ ID NO:39);
 CACCTGCCCCCTCART (SEQ ID NO:40); AGTCTCCTGGGAAYT (SEQ ID NO:41);
 CAGGAGACTCTTCYG (SEQ ID NO:42); TCAGAGTTACACCRG (SEQ ID NO:43);
 GCCCTTCTTCATCYT (SEQ ID NO:44); AGGACACTGGTGARG (SEQ ID NO:45);
 CTGATATACAGCAYG (SEQ ID NO:46); AGCCTTCAGCTCCRT (SEQ ID NO:47);
 45 TGATATACAGCACRG (SEQ ID NO:48); AAGCCTTCAGCTCYG (SEQ ID NO:49);
 CCTGAAGGATGGTYT (SEQ ID NO:50); and GTCCAGATGAGAARA (SEQ ID NO:51).

Other oligonucleotides of the invention hybridize to a target region located one to several nucleotides downstream of one of the novel polymorphic sites identified herein. Such oligonucleotides

are useful in polymerase-mediated primer extension methods for detecting one of the novel polymorphisms described herein and therefore such oligonucleotides are referred to herein as "primer-extension oligonucleotides". In a preferred embodiment, the 3'-terminus of a primer-extension oligonucleotide is a deoxynucleotide complementary to the nucleotide located immediately adjacent to the polymorphic site.

A particularly preferred oligonucleotide primer for detecting FY gene polymorphisms by primer extension terminates in a nucleotide sequence, listed 5' to 3', selected from the group consisting of:

10	ACTTGTCAGA	(SEQ ID NO:52);	GGAATACATG	(SEQ ID NO:53);
	CCAGACACCC	(SEQ ID NO:54);	AGGGGCTTGG	(SEQ ID NO:55);
	CCTCACATAC	(SEQ ID NO:56);	TGCACATATC	(SEQ ID NO:57);
	ACACAGCAAA	(SEQ ID NO:58);	CTCTGTGTAC	(SEQ ID NO:59);
	CTCACGCCCCA	(SEQ ID NO:60);	GTGTGTGCAC	(SEQ ID NO:61);
15	GGACAGAGTT	(SEQ ID NO:62);	TGGTGGTGGT	(SEQ ID NO:63);
	CACCACCTTT	(SEQ ID NO:64);	GTGTTTGGGA	(SEQ ID NO:65);
	CAATCTCCCT	(SEQ ID NO:66);	CCGAAGTGGA	(SEQ ID NO:67);
	TCTCCCTTCC	(SEQ ID NO:68);	GGAAAAAGC	(SEQ ID NO:69);
	CTTTTTCCTT	(SEQ ID NO:70);	AGGAAGGAGA	(SEQ ID NO:71);
20	CTGCCCTCA	(SEQ ID NO:72);	CTCCTGGGAA	(SEQ ID NO:73);
	GAGACTCTTC	(SEQ ID NO:74);	GAGTTACACC	(SEQ ID NO:75);
	CTTCTTCATC	(SEQ ID NO:76);	ACACTGGTGA	(SEQ ID NO:77);
	ATATACAGCA	(SEQ ID NO:78);	CTTCAGCTCC	(SEQ ID NO:79);
	TATACAGCAC	(SEQ ID NO:80);	CCTTCAGCTC	(SEQ ID NO:81);
25	GAAGGATGGT	(SEQ ID NO:82);	and CAGATGAGAA	(SEQ ID NO:83).

In some embodiments, a composition contains two or more differently labeled FY oligonucleotides for simultaneously probing the identity of nucleotides or nucleotide pairs at two or more polymorphic sites. It is also contemplated that primer compositions may contain two or more sets of allele-specific primer pairs to allow simultaneous targeting and amplification of two or more regions containing a polymorphic site.

FY oligonucleotides of the invention may also be immobilized on or synthesized on a solid surface such as a microchip, bead, or glass slide (see, e.g., WO 98/20020 and WO 98/20019). Such immobilized oligonucleotides may be used in a variety of polymorphism detection assays, including but not limited to probe hybridization and polymerase extension assays. Immobilized FY oligonucleotides of the invention may comprise an ordered array of oligonucleotides designed to rapidly screen a DNA sample for polymorphisms in multiple genes at the same time.

In another embodiment, the invention provides a kit comprising at least two FY oligonucleotides packaged in separate containers. The kit may also contain other components such as hybridization buffer (where the oligonucleotides are to be used as a probe) packaged in a separate container. Alternatively, where the oligonucleotides are to be used to amplify a target region, the kit may contain, packaged in separate containers, a polymerase and a reaction buffer optimized for primer extension mediated by the polymerase, such as PCR.

The above described oligonucleotide compositions and kits are useful in methods for genotyping and/or haplotyping the FY gene in an individual. As used herein, the terms "FY genotype" and "FY haplotype" mean the genotype or haplotype contains the nucleotide pair or nucleotide, respectively, that is present at one or more of the novel polymorphic sites described herein and may optionally also include the nucleotide pair or nucleotide present at one or more additional polymorphic sites in the FY gene. The additional polymorphic sites may be currently known polymorphic sites or sites that are subsequently discovered.

One embodiment of a genotyping method of the invention involves isolating from the individual a nucleic acid sample comprising the two copies of the FY gene, mRNA transcripts thereof or cDNA copies thereof, or a fragment of any of the foregoing, that are present in the individual, and determining the identity of the nucleotide pair at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS10, PS11, PS12, PS13, PS15, PS18, PS19 and PS20 in the two copies to assign a FY genotype to the individual. As will be readily understood by the skilled artisan, the two "copies" of a gene, mRNA or cDNA (or fragment of such FY molecules) in an individual may be the same allele or may be different alleles. In a preferred embodiment of the method for assigning a FY genotype, the identity of the nucleotide pair at one or more of the polymorphic sites selected from the group consisting of PS9, PS14, PS16 and PS17 is also determined. In another embodiment, a genotyping method of the invention comprises determining the identity of the nucleotide pair at each of PS1-PS20.

Typically, the nucleic acid sample is isolated from a biological sample taken from the individual, such as a blood sample or tissue sample. Suitable tissue samples include whole blood, semen, saliva, tears, urine, fecal material, sweat, buccal, skin and hair. The nucleic acid sample may be comprised of genomic DNA, mRNA, or cDNA and, in the latter two cases, the biological sample must be obtained from a tissue in which the FY gene is expressed. Furthermore it will be understood by the skilled artisan that mRNA or cDNA preparations would not be used to detect polymorphisms located in introns or in 5' and 3' untranslated regions if not present in the mRNA or cDNA. If a FY gene fragment is isolated, it must contain the polymorphic site(s) to be genotyped.

One embodiment of a haplotyping method of the invention comprises isolating from the individual a nucleic acid sample containing only one of the two copies of the FY gene, mRNA or cDNA, or a fragment of such FY molecules, that is present in the individual and determining in that copy the identity of the nucleotide at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS10, PS11, PS12, PS13, PS15, PS18, PS19 and PS20 in that copy to assign a FY haplotype to the individual.

The nucleic acid used in the above haplotyping methods of the invention may be isolated using any method capable of separating the two copies of the FY gene or fragment such as one of the methods described above for preparing FY isogenes, with targeted *in vivo* cloning being the preferred approach. As will be readily appreciated by those skilled in the art, any individual clone will typically

only provide haplotype information on one of the two FY gene copies present in an individual. If haplotype information is desired for the individual's other copy, additional FY clones will usually need to be examined. Typically, at least five clones should be examined to have more than a 90% probability of haplotyping both copies of the FY gene in an individual. In some cases, however, once the haplotype for one FY allele is directly determined, the haplotype for the other allele may be inferred if the individual has a known genotype for the polymorphic sites of interest or if the haplotype frequency or haplotype pair frequency for the individual's population group is known. In some embodiments, the FY haplotype is assigned to the individual by also identifying the nucleotide at one or more polymorphic sites selected from the group consisting of PS9, PS14, PS16 and PS17. In a particularly preferred embodiment, the nucleotide at each of PS1-PS20 is identified.

In another embodiment, the haplotyping method comprises determining whether an individual has one or more of the FY haplotypes shown in Table 4. This can be accomplished by identifying, for one or both copies of the individual's FY gene, the phased sequence of nucleotides present at each of PS1-PS20. This identifying step does not necessarily require that each of PS1-PS20 be directly examined. Typically only a subset of PS1-PS20 will need to be directly examined to assign to an individual one or more of the haplotypes shown in Table 4. This is because at least one polymorphic site in a gene is frequently in strong linkage disequilibrium with one or more other polymorphic sites in that gene (Drysdale, CM et al. 2000 *PNAS* 97:10483-10488; Rieder MJ et al. 1999 *Nature Genetics* 22:59-62). Two sites are said to be in linkage disequilibrium if the presence of a particular variant at one site enhances the predictability of another variant at the second site (Stephens, JC 1999, *Mol. Diag.* 4:309-317). Techniques for determining whether any two polymorphic sites are in linkage disequilibrium are well-known in the art (Weir B.S. 1996 *Genetic Data Analysis II*, Sinauer Associates, Inc. Publishers, Sunderland, MA).

In another embodiment of a haplotyping method of the invention, a FY haplotype pair is determined for an individual by identifying the phased sequence of nucleotides at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS10, PS11, PS12, PS13, PS15, PS18, PS19 and PS20 in each copy of the FY gene that is present in the individual. In a particularly preferred embodiment, the haplotyping method comprises identifying the phased sequence of nucleotides at each of PS1-PS20 in each copy of the FY gene.

When haplotyping both copies of the gene, the identifying step is preferably performed with each copy of the gene being placed in separate containers. However, it is also envisioned that if the two copies are labeled with different tags, or are otherwise separately distinguishable or identifiable, it could be possible in some cases to perform the method in the same container. For example, if first and second copies of the gene are labeled with different first and second fluorescent dyes, respectively, and an allele-specific oligonucleotide labeled with yet a third different fluorescent dye is used to assay the polymorphic site(s), then detecting a combination of the first and third dyes would identify the polymorphism in the first gene copy while detecting a combination of the second and third dyes would

identify the polymorphism in the second gene copy.

In both the genotyping and haplotyping methods, the identity of a nucleotide (or nucleotide pair) at a polymorphic site(s) may be determined by amplifying a target region(s) containing the polymorphic site(s) directly from one or both copies of the FY gene, or a fragment thereof, and the sequence of the amplified region(s) determined by conventional methods. It will be readily appreciated by the skilled artisan that only one nucleotide will be detected at a polymorphic site in individuals who are homozygous at that site, while two different nucleotides will be detected if the individual is heterozygous for that site. The polymorphism may be identified directly, known as positive-type identification, or by inference, referred to as negative-type identification. For example, where a SNP is known to be guanine and cytosine in a reference population, a site may be positively determined to be either guanine or cytosine for an individual homozygous at that site, or both guanine and cytosine, if the individual is heterozygous at that site. Alternatively, the site may be negatively determined to be not guanine (and thus cytosine/cytosine) or not cytosine (and thus guanine/guanine).

The target region(s) may be amplified using any oligonucleotide-directed amplification method, including but not limited to polymerase chain reaction (PCR) (U.S. Patent No. 4,965,188), ligase chain reaction (LCR) (Barany et al., *Proc. Natl. Acad. Sci. USA* 88:189-193, 1991; WO90/01069), and oligonucleotide ligation assay (OLA) (Landegren et al., *Science* 241:1077-1080, 1988). Other known nucleic acid amplification procedures may be used to amplify the target region including transcription-based amplification systems (U.S. Patent No. 5,130,238; EP 329,822; U.S. Patent No. 5,169,766, WO89/06700) and isothermal methods (Walker et al., *Proc. Natl. Acad. Sci. USA* 89:392-396, 1992).

A polymorphism in the target region may also be assayed before or after amplification using one of several hybridization-based methods known in the art. Typically, allele-specific oligonucleotides are utilized in performing such methods. The allele-specific oligonucleotides may be used as differently labeled probe pairs, with one member of the pair showing a perfect match to one variant of a target sequence and the other member showing a perfect match to a different variant. In some embodiments, more than one polymorphic site may be detected at once using a set of allele-specific oligonucleotides or oligonucleotide pairs. Preferably, the members of the set have melting temperatures within 5°C, and more preferably within 2°C, of each other when hybridizing to each of the polymorphic sites being detected.

Hybridization of an allele-specific oligonucleotide to a target polynucleotide may be performed with both entities in solution, or such hybridization may be performed when either the oligonucleotide or the target polynucleotide is covalently or noncovalently affixed to a solid support. Attachment may be mediated, for example, by antibody-antigen interactions, poly-L-Lys, streptavidin or avidin-biotin, salt bridges, hydrophobic interactions, chemical linkages, UV cross-linking baking, etc. Allele-specific oligonucleotides may be synthesized directly on the solid support or attached to the solid support subsequent to synthesis. Solid-supports suitable for use in detection methods of the invention

include substrates made of silicon, glass, plastic, paper and the like, which may be formed, for example, into wells (as in 96-well plates), slides, sheets, membranes, fibers, chips, dishes, and beads. The solid support may be treated, coated or derivatized to facilitate the immobilization of the allele-specific oligonucleotide or target nucleic acid.

5 The genotype or haplotype for the FY gene of an individual may also be determined by hybridization of a nucleic acid sample containing one or both copies of the gene, mRNA, cDNA or fragment(s) thereof, to nucleic acid arrays and subarrays such as described in WO 95/11995. The arrays would contain a battery of allele-specific oligonucleotides representing each of the polymorphic sites to be included in the genotype or haplotype.

10 The identity of polymorphisms may also be determined using a mismatch detection technique, including but not limited to the RNase protection method using riboprobes (Winter et al., *Proc. Natl. Acad. Sci. USA* 82:7575, 1985; Meyers et al., *Science* 230:1242, 1985) and proteins which recognize nucleotide mismatches, such as the E. coli mutS protein (Modrich, *P. Ann. Rev. Genet.* 25:229-253, 1991). Alternatively, variant alleles can be identified by single strand conformation polymorphism (SSCP) analysis (Orita et al., *Genomics* 5:874-879, 1989; Humphries et al., in *Molecular Diagnosis of Genetic Diseases*, R. Elles, ed., pp. 321-340, 1996) or denaturing gradient gel electrophoresis (DGGE) (Wartell et al., *Nucl. Acids Res.* 18:2699-2706, 1990; Sheffield et al., *Proc. Natl. Acad. Sci. USA* 86:232-236, 1989).

20 A polymerase-mediated primer extension method may also be used to identify the polymorphism(s). Several such methods have been described in the patent and scientific literature and include the "Genetic Bit Analysis" method (WO92/15712) and the ligase/polymerase mediated genetic bit analysis (U.S. Patent 5,679,524. Related methods are disclosed in WO91/02087, WO90/09455, WO95/17676, U.S. Patent Nos. 5,302,509, and 5,945,283. Extended primers containing a polymorphism may be detected by mass spectrometry as described in U.S. Patent No. 5,605,798.

25 Another primer extension method is allele-specific PCR (Ruano et al., *Nucl. Acids Res.* 17:8392, 1989; Ruano et al., *Nucl. Acids Res.* 19, 6877-6882, 1991; WO 93/22456; Turki et al., *J. Clin. Invest.* 95:1635-1641, 1995). In addition, multiple polymorphic sites may be investigated by simultaneously amplifying multiple regions of the nucleic acid using sets of allele-specific primers as described in Wallace et al. (WO89/10414).

30 In addition, the identity of the allele(s) present at any of the novel polymorphic sites described herein may be indirectly determined by haplotyping or genotyping another polymorphic site that is in linkage disequilibrium with the polymorphic site that is of interest. Polymorphic sites in linkage disequilibrium with the presently disclosed polymorphic sites may be located in regions of the gene or in other genomic regions not examined herein. Detection of the allele(s) present at a polymorphic site in linkage disequilibrium with the novel polymorphic sites described herein may be performed by, but
35 is not limited to, any of the above-mentioned methods for detecting the identity of the allele at a polymorphic site.

In another aspect of the invention, an individual's FY haplotype pair is predicted from its FY genotype using information on haplotype pairs known to exist in a reference population. In its broadest embodiment, the haplotyping prediction method comprises identifying a FY genotype for the individual at two or more FY polymorphic sites described herein, accessing data containing FY haplotype pairs identified in a reference population, and assigning a haplotype pair to the individual that is consistent with the genotype data. In one embodiment, the reference haplotype pairs include the FY haplotype pairs shown in Table 3. The FY haplotype pair can be assigned by comparing the individual's genotype with the genotypes corresponding to the haplotype pairs known to exist in the general population or in a specific population group, and determining which haplotype pair is consistent with the genotype of the individual. In some embodiments, the comparing step may be performed by visual inspection (for example, by consulting Table 3). When the genotype of the individual is consistent with more than one haplotype pair, frequency data (such as that presented in Table 6) may be used to determine which of these haplotype pairs is most likely to be present in the individual. This determination may also be performed in some embodiments by visual inspection, for example by consulting Table 6. If a particular FY haplotype pair consistent with the genotype of the individual is more frequent in the reference population than others consistent with the genotype, then that haplotype pair with the highest frequency is the most likely to be present in the individual. In other embodiments, the comparison may be made by a computer-implemented algorithm with the genotype of the individual and the reference haplotype data stored in computer-readable formats. For example, as described in PCT/US01/12831, filed April 18, 2001, one computer-implemented algorithm to perform this comparison entails enumerating all possible haplotype pairs which are consistent with the genotype, accessing data containing FY haplotype pairs frequency data determined in a reference population to determine a probability that the individual has a possible haplotype pair, and analyzing the determined probabilities to assign a haplotype pair to the individual.

Generally, the reference population should be composed of randomly-selected individuals representing the major ethnogeographic groups of the world. A preferred reference population for use in the methods of the present invention comprises an approximately equal number of individuals from Caucasian, African-descent, Asian and Hispanic-Latino population groups with the minimum number of each group being chosen based on how rare a haplotype one wants to be guaranteed to see. For example, if one wants to have a q% chance of not missing a haplotype that exists in the population at a p% frequency of occurring in the reference population, the number of individuals (n) who must be sampled is given by $2n = \log(1-q)/\log(1-p)$ where p and q are expressed as fractions. A preferred reference population allows the detection of any haplotype whose frequency is at least 10% with about 99% certainty and comprises about 20 unrelated individuals from each of the four population groups named above. A particularly preferred reference population includes a 3-generation family representing one or more of the four population groups to serve as controls for checking quality of haplotyping procedures.

In a preferred embodiment, the haplotype frequency data for each ethnogeographic group is examined to determine whether it is consistent with Hardy-Weinberg equilibrium. Hardy-Weinberg equilibrium (D.L. Hartl et al., Principles of Population Genomics, Sinauer Associates (Sunderland, MA), 3rd Ed., 1997) postulates that the frequency of finding the haplotype pair H_1 / H_2 is equal to

$$p_{H-W}(H_1 / H_2) = 2p(H_1)p(H_2) \text{ if } H_1 \neq H_2 \text{ and } p_{H-W}(H_1 / H_2) = p(H_1)p(H_2) \text{ if } H_1 = H_2.$$

A statistically significant difference between the observed and expected haplotype frequencies could be due to one or more factors including significant inbreeding in the population group, strong selective pressure on the gene, sampling bias, and/or errors in the genotyping process. If large deviations from Hardy-Weinberg equilibrium are observed in an ethnogeographic group, the number of individuals in that group can be increased to see if the deviation is due to a sampling bias. If a larger sample size does not reduce the difference between observed and expected haplotype pair frequencies, then one may wish to consider haplotyping the individual using a direct haplotyping method such as, for example, CLASPER System™ technology (U.S. Patent No. 5,866,404), single molecule dilution, or allele-specific long-range PCR (Michalotos-Beloin et al., *Nucleic Acids Res.* 24:4841-4843, 1996).

In one embodiment of this method for predicting a FY haplotype pair for an individual, the assigning step involves performing the following analysis. First, each of the possible haplotype pairs is compared to the haplotype pairs in the reference population. Generally, only one of the haplotype pairs in the reference population matches a possible haplotype pair and that pair is assigned to the individual. Occasionally, only one haplotype represented in the reference haplotype pairs is consistent with a possible haplotype pair for an individual, and in such cases the individual is assigned a haplotype pair containing this known haplotype and a new haplotype derived by subtracting the known haplotype from the possible haplotype pair. Alternatively, the haplotype pair in an individual may be predicted from the individual's genotype for that gene using reported methods (e.g., Clark et al. 1990 *Mol Bio Evol* 7:111-22; copending PCT/US01/12831 filed April 18, 2001) or through a commercial haplotyping service such as offered by Genaissance Pharmaceuticals, Inc. (New Haven, CT). In rare cases, either no haplotypes in the reference population are consistent with the possible haplotype pairs, or alternatively, multiple reference haplotype pairs are consistent with the possible haplotype pairs. In such cases, the individual is preferably haplotyped using a direct molecular haplotyping method such as, for example, CLASPER System™ technology (U.S. Patent No. 5,866,404), SMD, or allele-specific long-range PCR (Michalotos-Beloin et al., *supra*).

The invention also provides a method for determining the frequency of a FY genotype, haplotype, or haplotype pair in a population. The method comprises, for each member of the population, determining the genotype or the haplotype pair for the novel FY polymorphic sites described herein, and calculating the frequency any particular genotype, haplotype, or haplotype pair is found in the population. The population may be e.g., a reference population, a family population, a same gender population, a population group, or a trait population (e.g., a group of individuals exhibiting a trait of interest such as a medical condition or response to a therapeutic treatment).

In another aspect of the invention, frequency data for FY genotypes, haplotypes, and/or haplotype pairs are determined in a reference population and used in a method for identifying an association between a trait and a FY genotype, haplotype, or haplotype pair. The trait may be any detectable phenotype, including but not limited to susceptibility to a disease or response to a treatment.

5 In one embodiment, the method involves obtaining data on the frequency of the genotype(s), haplotype(s), or haplotype pair(s) of interest in a reference population as well as in a population exhibiting the trait. Frequency data for one or both of the reference and trait populations may be obtained by genotyping or haplotyping each individual in the populations using one or more of the methods described above. The haplotypes for the trait population may be determined directly or,
10 alternatively, by a predictive genotype to haplotype approach as described above. In another embodiment, the frequency data for the reference and/or trait populations is obtained by accessing previously determined frequency data, which may be in written or electronic form. For example, the frequency data may be present in a database that is accessible by a computer. Once the frequency data is obtained, the frequencies of the genotype(s), haplotype(s), or haplotype pair(s) of interest in the
15 reference and trait populations are compared. In a preferred embodiment, the frequencies of all genotypes, haplotypes, and/or haplotype pairs observed in the populations are compared. If a particular FY genotype, haplotype, or haplotype pair is more frequent in the trait population than in the reference population at a statistically significant amount, then the trait is predicted to be associated with that FY genotype, haplotype or haplotype pair. Preferably, the FY genotype, haplotype, or
20 haplotype pair being compared in the trait and reference populations is selected from the full-genotypes and full-haplotypes shown in Tables 3 and 4, or from sub-genotypes and sub-haplotypes derived from these genotypes and haplotypes. Sub-genotypes useful in the invention preferably do not include sub-genotypes solely for any one of PS9, PS14, PS16 and PS17 or for any combination thereof.

25 In a preferred embodiment of the method, the trait of interest is a clinical response exhibited by a patient to some therapeutic treatment, for example, response to a drug targeting FY or response to a therapeutic treatment for a medical condition. As used herein, "medical condition" includes but is not limited to any condition or disease manifested as one or more physical and/or psychological symptoms for which treatment is desirable, and includes previously and newly identified diseases and
30 other disorders. As used herein the term "clinical response" means any or all of the following: a quantitative measure of the response, no response, and/or adverse response (i.e., side effects).

In order to deduce a correlation between clinical response to a treatment and a FY genotype, haplotype, or haplotype pair, it is necessary to obtain data on the clinical responses exhibited by a population of individuals who received the treatment, hereinafter the "clinical population". This
35 clinical data may be obtained by analyzing the results of a clinical trial that has already been run and/or the clinical data may be obtained by designing and carrying out one or more new clinical trials. As used herein, the term "clinical trial" means any research study designed to collect clinical data on

responses to a particular treatment, and includes but is not limited to phase I, phase II and phase III clinical trials. Standard methods are used to define the patient population and to enroll subjects.

It is preferred that the individuals included in the clinical population have been graded for the existence of the medical condition of interest. This is important in cases where the symptom(s) being presented by the patients can be caused by more than one underlying condition, and where treatment of the underlying conditions are not the same. An example of this would be where patients experience breathing difficulties that are due to either asthma or respiratory infections. If both sets were treated with an asthma medication, there would be a spurious group of apparent non-responders that did not actually have asthma. These people would affect the ability to detect any correlation between haplotype and treatment outcome. This grading of potential patients could employ a standard physical exam or one or more lab tests. Alternatively, grading of patients could use haplotyping for situations where there is a strong correlation between haplotype pair and disease susceptibility or severity.

The therapeutic treatment of interest is administered to each individual in the trial population and each individual's response to the treatment is measured using one or more predetermined criteria. It is contemplated that in many cases, the trial population will exhibit a range of responses and that the investigator will choose the number of responder groups (e.g., low, medium, high) made up by the various responses. In addition, the FY gene for each individual in the trial population is genotyped and/or haplotyped, which may be done before or after administering the treatment.

After both the clinical and polymorphism data have been obtained, correlations between individual response and FY genotype or haplotype content are created. Correlations may be produced in several ways. In one method, individuals are grouped by their FY genotype or haplotype (or haplotype pair) (also referred to as a polymorphism group), and then the averages and standard deviations of clinical responses exhibited by the members of each polymorphism group are calculated.

These results are then analyzed to determine if any observed variation in clinical response between polymorphism groups is statistically significant. Statistical analysis methods which may be used are described in L.D. Fisher and G. vanBelle, "Biostatistics: A Methodology for the Health Sciences", Wiley-Interscience (New York) 1993. This analysis may also include a regression calculation of which polymorphic sites in the FY gene give the most significant contribution to the differences in phenotype. One regression model useful in the invention is described in WO 01/01218, entitled "Methods for Obtaining and Using Haplotype Data".

A second method for finding correlations between FY haplotype content and clinical responses uses predictive models based on error-minimizing optimization algorithms. One of many possible optimization algorithms is a genetic algorithm (R. Judson, "Genetic Algorithms and Their Uses in Chemistry" in Reviews in Computational Chemistry, Vol. 10, pp. 1-73, K. B. Lipkowitz and D. B. Boyd, eds. (VCH Publishers, New York, 1997). Simulated annealing (Press et al., "Numerical Recipes in C: The Art of Scientific Computing", Cambridge University Press (Cambridge) 1992, Ch. 10), neural networks (E. Rich and K. Knight, "Artificial Intelligence", 2nd Edition (McGraw-Hill, New

York, 1991, Ch. 18), standard gradient descent methods (Press et al., *supra*, Ch. 10), or other global or local optimization approaches (see discussion in Judson, *supra*) could also be used. Preferably, the correlation is found using a genetic algorithm approach as described in WO 01/01218.

Correlations may also be analyzed using analysis of variation (ANOVA) techniques to
5 determine how much of the variation in the clinical data is explained by different subsets of the polymorphic sites in the FY gene. As described in WO 01/01218, ANOVA is used to test hypotheses about whether a response variable is caused by or correlated with one or more traits or variables that can be measured (Fisher and vanBelle, *supra*, Ch. 10).

From the analyses described above, a mathematical model may be readily constructed by the
10 skilled artisan that predicts clinical response as a function of FY genotype or haplotype content. Preferably, the model is validated in one or more follow-up clinical trials designed to test the model.

The identification of an association between a clinical response and a genotype or haplotype (or haplotype pair) for the FY gene may be the basis for designing a diagnostic method to determine those individuals who will or will not respond to the treatment, or alternatively, will respond at a lower
15 level and thus may require more treatment, i.e., a greater dose of a drug. The diagnostic method may take one of several forms: for example, a direct DNA test (i.e., genotyping or haplotyping one or more of the polymorphic sites in the FY gene), a serological test, or a physical exam measurement. The only requirement is that there be a good correlation between the diagnostic test results and the underlying FY genotype or haplotype that is in turn correlated with the clinical response. In a preferred
20 embodiment, this diagnostic method uses the predictive haplotyping method described above.

In another embodiment, the invention provides an isolated polynucleotide comprising a polymorphic variant of the FY gene or a fragment of the gene which contains at least one of the novel polymorphic sites described herein. The nucleotide sequence of a variant FY gene is identical to the reference genomic sequence for those portions of the gene examined, as described in the Examples
25 below, except that it comprises a different nucleotide at one or more of the novel polymorphic sites PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS10, PS11, PS12, PS13, PS15, PS18, PS19 and PS20, and may also comprise one or more additional polymorphisms selected from the group consisting of cytosine at PS9, guanine at PS14, thymine at PS16 and adenine at PS17. Similarly, the nucleotide sequence of a variant fragment of the FY gene is identical to the corresponding portion of the reference
30 sequence except for having a different nucleotide at one or more of the novel polymorphic sites described herein. Thus, the invention specifically does not include polynucleotides comprising a nucleotide sequence identical to the reference sequence of the FY gene (or other reported FY sequences) or to portions of the reference sequence (or other reported FY sequences), except for the haplotyping and genotyping oligonucleotides described above.

35 The location of a polymorphism in a variant FY gene or fragment is preferably identified by aligning its sequence against SEQ ID NO:1. The polymorphism is selected from the group consisting of thymine at PS1, adenine at PS2, guanine at PS3, thymine at PS4, adenine at PS5, cytosine at PS6,

thymine at PS7, cytosine at PS8, thymine at PS10, thymine at PS11, guanine at PS12, thymine at PS13, thymine at PS15, thymine at PS18, adenine at PS19 and thymine at PS20. In a preferred embodiment, the polymorphic variant comprises a naturally-occurring isogene of the FY gene which is defined by any one of haplotypes 1-23 shown in Table 4 below.

5 Polymorphic variants of the invention may be prepared by isolating a clone containing the FY gene from a human genomic library. The clone may be sequenced to determine the identity of the nucleotides at the novel polymorphic sites described herein. Any particular variant or fragment thereof, that is claimed herein could be prepared from this clone by performing *in vitro* mutagenesis using procedures well-known in the art. Any particular FY variant or fragment thereof may also be
10 prepared using synthetic or semi-synthetic methods known in the art.

FY isogenes, or fragments thereof, may be isolated using any method that allows separation of the two "copies" of the FY gene present in an individual, which, as readily understood by the skilled artisan, may be the same allele or different alleles. Separation methods include targeted *in vivo* cloning (TIVC) in yeast as described in WO 98/01573, U.S. Patent No. 5,866,404, and U.S. Patent No.
15 5,972,614. Another method, which is described in U.S. Patent No. 5,972,614, uses an allele specific oligonucleotide in combination with primer extension and exonuclease degradation to generate hemizygous DNA targets. Yet other methods are single molecule dilution (SMD) as described in Ruaño et al., *Proc. Natl. Acad. Sci.* 87:6296-6300, 1990; and allele specific PCR (Ruaño et al., 1989, *supra*; Ruaño et al., 1991, *supra*; Michalatos-Beloin et al., *supra*).

20 The invention also provides FY genome anthologies, which are collections of at least two FY isogenes found in a given population. The population may be any group of at least two individuals, including but not limited to a reference population, a population group, a family population, a clinical population, and a same gender population. A FY genome anthology may comprise individual FY isogenes stored in separate containers such as microtest tubes, separate wells of a microtitre plate and
25 the like. Alternatively, two or more groups of the FY isogenes in the anthology may be stored in separate containers. Individual isogenes or groups of such isogenes in a genome anthology may be stored in any convenient and stable form, including but not limited to in buffered solutions, as DNA precipitates, freeze-dried preparations and the like. A preferred FY genome anthology of the invention comprises a set of isogenes defined by the haplotypes shown in Table 4 below.

30 An isolated polynucleotide containing a polymorphic variant nucleotide sequence of the invention may be operably linked to one or more expression regulatory elements in a recombinant expression vector capable of being propagated and expressing the encoded FY protein in a prokaryotic or a eukaryotic host cell. Examples of expression regulatory elements which may be used include, but are not limited to, the lac system, operator and promoter regions of phage lambda, yeast promoters, and
35 promoters derived from vaccinia virus, adenovirus, retroviruses, or SV40. Other regulatory elements include, but are not limited to, appropriate leader sequences, termination codons, polyadenylation signals, and other sequences required for the appropriate transcription and subsequent translation of the

nucleic acid sequence in a given host cell. Of course, the correct combinations of expression regulatory elements will depend on the host system used. In addition, it is understood that the expression vector contains any additional elements necessary for its transfer to and subsequent replication in the host cell. Examples of such elements include, but are not limited to, origins of replication and selectable markers. Such expression vectors are commercially available or are readily constructed using methods known to those in the art (e.g., F. Ausubel et al., 1987, in "Current Protocols in Molecular Biology", John Wiley and Sons, New York, New York). Host cells which may be used to express the variant FY sequences of the invention include, but are not limited to, eukaryotic and mammalian cells, such as animal, plant, insect and yeast cells, and prokaryotic cells, such as E. coli, or algal cells as known in the art. The recombinant expression vector may be introduced into the host cell using any method known to those in the art including, but not limited to, microinjection, electroporation, particle bombardment, transduction, and transfection using DEAE-dextran, lipofection, or calcium phosphate (see e.g., Sambrook et al. (1989) in "Molecular Cloning. A Laboratory Manual", Cold Spring Harbor Press, Plainview, New York). In a preferred aspect, eukaryotic expression vectors that function in eukaryotic cells, and preferably mammalian cells, are used. Non-limiting examples of such vectors include vaccinia virus vectors, adenovirus vectors, herpes virus vectors, and baculovirus transfer vectors. Preferred eukaryotic cell lines include COS cells, CHO cells, HeLa cells, NIH/3T3 cells, and embryonic stem cells (Thomson, J. A. et al., 1998 *Science* 282:1145-1147). Particularly preferred host cells are mammalian cells.

As will be readily recognized by the skilled artisan, expression of polymorphic variants of the FY gene will produce FY mRNAs varying from each other at any polymorphic site retained in the spliced and processed mRNA molecules. These mRNAs can be used for the preparation of a FY cDNA comprising a nucleotide sequence which is a polymorphic variant of the FY reference coding sequence shown in Figure 2. Thus, the invention also provides FY mRNAs and corresponding cDNAs which comprise a nucleotide sequence that is identical to SEQ ID NO:2 (Fig. 2) (or its corresponding RNA sequence) for those regions of SEQ ID NO:2 that correspond to the examined portions of the FY gene (as described in the Examples below), except for having one or more polymorphisms selected from the group consisting of thymine at a position corresponding to nucleotide 205, thymine at a position corresponding to nucleotide 608, adenine at a position corresponding to nucleotide 609 and thymine at a position corresponding to nucleotide 983, and may also comprise one or more additional polymorphisms selected from the group consisting of guanine at a position corresponding to nucleotide 131, thymine at a position corresponding to nucleotide 271 and adenine at a position corresponding to nucleotide 304. A particularly preferred polymorphic cDNA variant comprises the coding sequence of a FY isogene defined by any one of haplotypes 2-4, 6, 9, 10-12, and 16-18. Fragments of these variant mRNAs and cDNAs are included in the scope of the invention, provided they contain one or more of the novel polymorphisms described herein. The invention specifically excludes polynucleotides identical to previously identified FY mRNAs or cDNAs, and previously described fragments thereof.

Polynucleotides comprising a variant FY RNA or DNA sequence may be isolated from a biological sample using well-known molecular biological procedures or may be chemically synthesized.

As used herein, a polymorphic variant of a FY gene, mRNA or cDNA fragment comprises at least one novel polymorphism identified herein and has a length of at least 10 nucleotides and may range up to the full length of the gene. Preferably, such fragments are between 100 and 3000 nucleotides in length, and more preferably between 200 and 2000 nucleotides in length, and most preferably between 500 and 1000 nucleotides in length.

In describing the FY polymorphic sites identified herein, reference is made to the sense strand of the gene for convenience. However, as recognized by the skilled artisan, nucleic acid molecules containing the FY gene or cDNA may be complementary double stranded molecules and thus reference to a particular site on the sense strand refers as well to the corresponding site on the complementary antisense strand. Thus, reference may be made to the same polymorphic site on either strand and an oligonucleotide may be designed to hybridize specifically to either strand at a target region containing the polymorphic site. Thus, the invention also includes single-stranded polynucleotides which are complementary to the sense strand of the FY genomic, mRNA and cDNA variants described herein.

Polynucleotides comprising a polymorphic gene variant or fragment of the invention may be useful for therapeutic purposes. For example, where a patient could benefit from expression, or increased expression, of a particular FY protein isoform, an expression vector encoding the isoform may be administered to the patient. The patient may be one who lacks the FY isogene encoding that isoform or may already have at least one copy of that isogene.

In other situations, it may be desirable to decrease or block expression of a particular FY isogene. Expression of a FY isogene may be turned off by transforming a targeted organ, tissue or cell population with an expression vector that expresses high levels of untranslatable mRNA or antisense RNA for the isogene or fragment thereof. Alternatively, oligonucleotides directed against the regulatory regions (e.g., promoter, introns, enhancers, 3' untranslated region) of the isogene may block transcription. Oligonucleotides targeting the transcription initiation site, e.g., between positions -10 and +10 from the start site are preferred. Similarly, inhibition of transcription can be achieved using oligonucleotides that base-pair with region(s) of the isogene DNA to form triplex DNA (see e.g., Gee et al. in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, N.Y., 1994). Antisense oligonucleotides may also be designed to block translation of FY mRNA transcribed from a particular isogene. It is also contemplated that ribozymes may be designed that can catalyze the specific cleavage of FY mRNA transcribed from a particular isogene.

The untranslated mRNA, antisense RNA or antisense oligonucleotides may be delivered to a target cell or tissue by expression from a vector introduced into the cell or tissue *in vivo* or *ex vivo*. Alternatively, such molecules may be formulated as a pharmaceutical composition for administration to the patient. Oligoribonucleotides and/or oligodeoxynucleotides intended for use as antisense

oligonucleotides may be modified to increase stability and half-life. Possible modifications include, but are not limited to phosphorothioate or 2' O-methyl linkages, and the inclusion of nontraditional bases such as inosine and queosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytosine, guanine, thymine, and uracil which are not as easily recognized by endogenous nucleases.

The invention also provides an isolated polypeptide comprising a polymorphic variant of (a) the reference FY amino acid sequence shown in Figure 3 or (b) a fragment of this reference sequence. The location of a variant amino acid in a FY polypeptide or fragment of the invention is preferably identified by aligning its sequence against SEQ ID NO:3 (Fig. 3). A FY protein variant of the invention comprises an amino acid sequence identical to SEQ ID NO:3 for those regions of SEQ ID NO:3 that are encoded by examined portions of the FY gene (as described in the Examples below), except for having one or more variant amino acids selected from the group consisting of phenylalanine at a position corresponding to amino acid position 69, isoleucine at a position corresponding to amino acid position 203, isoleucine at a position corresponding to amino acid position 203 and phenylalanine at a position corresponding to amino acid position 328, and may also comprise one or more additional variant amino acids selected from the group consisting of glycine at a position corresponding to amino acid position 44, cysteine at a position corresponding to amino acid position 91 and threonine at a position corresponding to amino acid position 102. Thus, a FY fragment of the invention, also referred to herein as a FY peptide variant, is any fragment of a FY protein variant that contains one or more of the amino acid variations described herein. The invention specifically excludes amino acid sequences identical to those previously identified for FY, including SEQ ID NO:3, and previously described fragments thereof. FY protein variants included within the invention comprise all amino acid sequences based on SEQ ID NO:3 and having any combination of amino acid variations described herein. In preferred embodiments, a FY protein variant of the invention is encoded by an isogene defined by one of the observed haplotypes, 2-4, 6, 9, 10-12, and 16-18, shown in Table 4.

A FY peptide variant of the invention is at least 6 amino acids in length and is preferably any number between 6 and 30 amino acids long, more preferably between 10 and 25, and most preferably between 15 and 20 amino acids long. Such FY peptide variants may be useful as antigens to generate antibodies specific for one of the above FY isoforms. In addition, the FY peptide variants may be useful in drug screening assays.

A FY variant protein or peptide of the invention may be prepared by chemical synthesis or by expressing an appropriate variant FY genomic or cDNA sequence described above. Alternatively, the FY protein variant may be isolated from a biological sample of an individual having a FY isogene which encodes the variant protein. Where the sample contains two different FY isoforms (i.e., the individual has different FY isogenes), a particular FY isoform of the invention can be isolated by immunoaffinity chromatography using an antibody which specifically binds to that particular FY isoform but does not bind to the other FY isoform.

----- The expressed or isolated FY protein or peptide may be detected by methods known in the art, including Coomassie blue staining, silver staining, and Western blot analysis using antibodies specific for the isoform of the FY protein or peptide as discussed further below. FY variant proteins and peptides can be purified by standard protein purification procedures known in the art, including
5 differential precipitation, molecular sieve chromatography, ion-exchange chromatography, isoelectric focusing, gel electrophoresis, affinity and immunoaffinity chromatography and the like. (Ausubel et. al., 1987, In Current Protocols in Molecular Biology John Wiley and Sons, New York, New York). In the case of immunoaffinity chromatography, antibodies specific for a particular polymorphic variant may be used.

10 A polymorphic variant FY gene of the invention may also be fused in frame with a heterologous sequence to encode a chimeric FY protein. The non-FY portion of the chimeric protein may be recognized by a commercially available antibody. In addition, the chimeric protein may also be engineered to contain a cleavage site located between the FY and non-FY portions so that the FY protein may be cleaved and purified away from the non-FY portion.

15 An additional embodiment of the invention relates to using a novel FY protein isoform, or a fragment thereof, in any of a variety of drug screening assays. Such screening assays may be performed to identify agents that bind specifically to all known FY protein isoforms or to only a subset of one or more of these isoforms. The agents may be from chemical compound libraries, peptide libraries and the like. The FY protein or peptide variant may be free in solution or affixed to a solid
20 support. In one embodiment, high throughput screening of compounds for binding to a FY variant may be accomplished using the method described in PCT application WO84/03565, in which large numbers of test compounds are synthesized on a solid substrate, such as plastic pins or some other surface, contacted with the FY protein(s) of interest and then washed. Bound FY protein(s) are then detected using methods well-known in the art.

25 In another embodiment, a novel FY protein isoform may be used in assays to measure the binding affinities of one or more candidate drugs targeting the FY protein.

In yet another embodiment, when a particular FY haplotype or group of FY haplotypes encodes a FY protein variant with an amino acid sequence distinct from that of FY protein isoforms encoded by other FY haplotypes, then detection of that particular FY haplotype or group of FY
30 haplotypes may be accomplished by detecting expression of the encoded FY protein variant using any of the methods described herein or otherwise commonly known to the skilled artisan.

In another embodiment, the invention provides antibodies specific for and immunoreactive with one or more of the novel FY protein or peptide variants described herein. The antibodies may be either monoclonal or polyclonal in origin. The FY protein or peptide variant used to generate the
35 antibodies may be from natural or recombinant sources (in vitro or in vivo) or produced by chemical synthesis or semi-synthetic synthesis using synthesis techniques known in the art. If the FY protein or peptide variant is of insufficient size to be antigenic, it may be concatenated or conjugated, complexed,

or otherwise covalently linked to a carrier molecule to enhance the antigenicity of the peptide. Examples of carrier molecules, include, but are not limited to, albumins (e.g., human, bovine, fish, ovine), and keyhole limpet hemocyanin (Basic and Clinical Immunology, 1991, Eds. D.P. Stites, and A.I. Terr, Appleton and Lange, Norwalk Connecticut, San Mateo, California).

5 In one embodiment, an antibody specifically immunoreactive with one of the novel protein or peptide variants described herein is administered to an individual to neutralize activity of the FY isoform expressed by that individual. The antibody may be formulated as a pharmaceutical composition which includes a pharmaceutically acceptable carrier.

10 Antibodies specific for and immunoreactive with one of the novel protein isoforms described herein may be used to immunoprecipitate the FY protein variant from solution as well as react with FY protein isoforms on Western or immunoblots of polyacrylamide gels on membrane supports or substrates. In another preferred embodiment, the antibodies will detect FY protein isoforms in paraffin or frozen tissue sections, or in cells which have been fixed or unfixed and prepared on slides, coverslips, or the like, for use in immunocytochemical, immunohistochemical, and
15 immunofluorescence techniques.

In another embodiment, an antibody specifically immunoreactive with one of the novel FY protein variants described herein is used in immunoassays to detect this variant in biological samples. In this method, an antibody of the present invention is contacted with a biological sample and the formation of a complex between the FY protein variant and the antibody is detected. As described,
20 suitable immunoassays include radioimmunoassay, Western blot assay, immunofluorescent assay, enzyme linked immunoassay (ELISA), chemiluminescent assay, immunohistochemical assay, immunocytochemical assay, and the like (see, e.g., Principles and Practice of Immunoassay, 1991, Eds. Christopher P. Price and David J. Neoman, Stockton Press, New York, New York; Current Protocols in Molecular Biology, 1987, Eds. Ausubel et al., John Wiley and Sons, New York, New York).
25 Standard techniques known in the art for ELISA are described in Methods in Immunodiagnosis, 2nd Ed., Eds. Rose and Bigazzi, John Wiley and Sons, New York 1980; and Campbell et al., 1984, Methods in Immunology, W.A. Benjamin, Inc.). Such assays may be direct, indirect, competitive, or noncompetitive as described in the art (see, e.g., Principles and Practice of Immunoassay, 1991, Eds. Christopher P. Price and David J. Neoman, Stockton Pres, NY, NY; and Oellirich, M., 1984, J. Clin.
30 Chem. Clin. Biochem., 22:895-904). Proteins may be isolated from test specimens and biological samples by conventional methods, as described in Current Protocols in Molecular Biology, supra.

Exemplary antibody molecules for use in the detection and therapy methods of the present invention are intact immunoglobulin molecules, substantially intact immunoglobulin molecules, or those portions of immunoglobulin molecules that contain the antigen binding site. Polyclonal or
35 monoclonal antibodies may be produced by methods conventionally known in the art (e.g., Kohler and Milstein, 1975, Nature, 256:495-497; Campbell Monoclonal Antibody Technology, the Production and Characterization of Rodent and Human Hybridomas, 1985, In: Laboratory Techniques in Biochemistry

and Molecular Biology, Eds. Burdon et al., Volume 13, Elsevier Science Publishers, Amsterdam). The antibodies or antigen binding fragments thereof may also be produced by genetic engineering. The technology for expression of both heavy and light chain genes in *E. coli* is the subject of PCT patent applications, publication number WO 901443, WO 901443 and WO 9014424 and in Huse et al., 1989, Science, 246:1275-1281. The antibodies may also be humanized (e.g., Queen, C. et al. 1989 Proc. Natl. Acad. Sci. USA 86:10029).

Effect(s) of the polymorphisms identified herein on expression of FY may be investigated by various means known in the art, such as by *in vitro* translation of mRNA transcripts of the FY gene, cDNA or fragment thereof, or by preparing recombinant cells and/or nonhuman recombinant organisms, preferably recombinant animals, containing a polymorphic variant of the FY gene. As used herein, "expression" includes but is not limited to one or more of the following: transcription of the gene into precursor mRNA; splicing and other processing of the precursor mRNA to produce mature mRNA; mRNA stability; translation of the mature mRNA(s) into FY protein(s) (including effects of polymorphisms on codon usage and tRNA availability); and glycosylation and/or other modifications of the translation product, if required for proper expression and function.

To prepare a recombinant cell of the invention, the desired FY isogene, cDNA or coding sequence may be introduced into the cell in a vector such that the isogene, cDNA or coding sequence remains extrachromosomal. In such a situation, the gene will be expressed by the cell from the extrachromosomal location. In a preferred embodiment, the FY isogene, cDNA or coding sequence is introduced into a cell in such a way that it recombines with the endogenous FY gene present in the cell. Such recombination requires the occurrence of a double recombination event, thereby resulting in the desired FY gene polymorphism. Vectors for the introduction of genes both for recombination and for extrachromosomal maintenance are known in the art, and any suitable vector or vector construct may be used in the invention. Methods such as electroporation, particle bombardment, calcium phosphate co-precipitation and viral transduction for introducing DNA into cells are known in the art; therefore, the choice of method may lie with the competence and preference of the skilled practitioner. Examples of cells into which the FY isogene, cDNA or coding sequence may be introduced include, but are not limited to, continuous culture cells, such as COS, CHO, NIH/3T3, and primary or culture cells of the relevant tissue type, i.e., they express the FY isogene, cDNA or coding sequence. Such recombinant cells can be used to compare the biological activities of the different protein variants.

Recombinant nonhuman organisms, i.e., transgenic animals, expressing a variant FY gene, cDNA or coding sequence are prepared using standard procedures known in the art. Preferably, a construct comprising the variant gene, cDNA or coding sequence is introduced into a nonhuman animal or an ancestor of the animal at an embryonic stage, i.e., the one-cell stage, or generally not later than about the eight-cell stage. Transgenic animals carrying the constructs of the invention can be made by several methods known to those having skill in the art. One method involves transfecting into the embryo a retrovirus constructed to contain one or more insulator elements, a gene or genes (or

cDNA or coding sequence) of interest, and other components known to those skilled in the art to provide a complete shuttle vector harboring the insulated gene(s) as a transgene, see e.g., U.S. Patent No. 5,610,053. Another method involves directly injecting a transgene into the embryo. A third method involves the use of embryonic stem cells. Examples of animals into which the FY isogene, cDNA or coding sequences may be introduced include, but are not limited to, mice, rats, other rodents, and nonhuman primates (see "The Introduction of Foreign Genes into Mice" and the cited references therein, In: Recombinant DNA, Eds. J.D. Watson, M. Gilman, J. Witkowski, and M. Zoller; W.H. Freeman and Company, New York, pages 254-272). Transgenic animals stably expressing a human FY isogene, cDNA or coding sequence and producing the encoded human FY protein can be used as biological models for studying diseases related to abnormal FY expression and/or activity, and for screening and assaying various candidate drugs, compounds, and treatment regimens to reduce the symptoms or effects of these diseases.

An additional embodiment of the invention relates to pharmaceutical compositions for treating disorders affected by expression or function of a novel FY isogene described herein. The pharmaceutical composition may comprise any of the following active ingredients: a polynucleotide comprising one of these novel FY isogenes (or cDNAs or coding sequences); an antisense oligonucleotide directed against one of the novel FY isogenes, a polynucleotide encoding such an antisense oligonucleotide, or another compound which inhibits expression of a novel FY isogene described herein. Preferably, the composition contains the active ingredient in a therapeutically effective amount. By therapeutically effective amount is meant that one or more of the symptoms relating to disorders affected by expression or function of a novel FY isogene is reduced and/or eliminated. The composition also comprises a pharmaceutically acceptable carrier, examples of which include, but are not limited to, saline, buffered saline, dextrose, and water. Those skilled in the art may employ a formulation most suitable for the active ingredient, whether it is a polynucleotide, oligonucleotide, protein, peptide or small molecule antagonist. The pharmaceutical composition may be administered alone or in combination with at least one other agent, such as a stabilizing compound. Administration of the pharmaceutical composition may be by any number of routes including, but not limited to oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, intradermal, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, PA).

For any composition, determination of the therapeutically effective dose of active ingredient and/or the appropriate route of administration is well within the capability of those skilled in the art. For example, the dose can be estimated initially either in cell culture assays or in animal models. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans. The exact dosage will be determined by the practitioner, in light of factors

relating to the patient requiring treatment, including but not limited to severity of the disease state, general health, age, weight and gender of the patient, diet, time and frequency of administration, other drugs being taken by the patient, and tolerance/response to the treatment.

Any or all analytical and mathematical operations involved in practicing the methods of the present invention may be implemented by a computer. In addition, the computer may execute a program that generates views (or screens) displayed on a display device and with which the user can interact to view and analyze large amounts of information relating to the FY gene and its genomic variation, including chromosome location, gene structure, and gene family, gene expression data, polymorphism data, genetic sequence data, and clinical data population data (e.g., data on ethnogeographic origin, clinical responses, genotypes, and haplotypes for one or more populations). The FY polymorphism data described herein may be stored as part of a relational database (e.g., an instance of an Oracle database or a set of ASCII flat files). These polymorphism data may be stored on the computer's hard drive or may, for example, be stored on a CD-ROM or on one or more other storage devices accessible by the computer. For example, the data may be stored on one or more databases in communication with the computer via a network.

Preferred embodiments of the invention are described in the following examples. Other embodiments within the scope of the claims herein will be apparent to one skilled in the art from consideration of the specification or practice of the invention as disclosed herein. It is intended that the specification, together with the examples, be considered exemplary only, with the scope and spirit of the invention being indicated by the claims which follow the examples.

EXAMPLES

The Examples herein are meant to exemplify the various aspects of carrying out the invention and are not intended to limit the scope of the invention in any way. The Examples do not include detailed descriptions for conventional methods employed, such as in the performance of genomic DNA isolation, PCR and sequencing procedures. Such methods are well-known to those skilled in the art and are described in numerous publications, for example, Sambrook, Fritsch, and Maniatis, "Molecular Cloning: A Laboratory Manual", 2nd Edition, Cold Spring Harbor Laboratory Press, USA, (1989).

EXAMPLE 1

This example illustrates examination of various regions of the FY gene for polymorphic sites.

Amplification of Target Regions

The following target regions of the FY gene were amplified using PCR primer pairs. The primers used for each region are represented below by providing the nucleotide positions of their initial and final nucleotides, which correspond to positions in SEQ ID NO:1 (Figure 1).

PCR Primer Pairs

Fragment	Forward Primer	Reverse Primer	PCR Product
Fragment 1	2486 - 2508	complement of 3095 - 3074	610 nt
Fragment 2	2773 - 2795	complement of 3355 - 3333	583 nt
Fragment 3	2812 - 2834	complement of 3318 - 3295	507 nt
Fragment 4	3048 - 3073	complement of 3562 - 3541	515 nt
Fragment 5	3255 - 3278	complement of 3779 - 3757	525 nt
Fragment 6	3872 - 3892	complement of 4431 - 4409	560 nt
Fragment 7	4118 - 4140	complement of 4726 - 4705	609 nt
Fragment 8	4402 - 4423	complement of 5039 - 5016	638 nt
Fragment 9	4706 - 4727	complement of 5302 - 5280	597 nt

These primer pairs were used in PCR reactions containing genomic DNA isolated from immortalized cell lines for each member of the Index Repository. The PCR reactions were carried out under the following conditions:

Reaction volume	= 10 μ l
10 x Advantage 2 Polymerase reaction buffer (Clontech)	= 1 μ l
100 ng of human genomic DNA	= 1 μ l
10 mM dNTP	= 0.4 μ l
Advantage 2 Polymerase enzyme mix (Clontech)	= 0.2 μ l
Forward Primer (10 μ M)	= 0.4 μ l
Reverse Primer (10 μ M)	= 0.4 μ l
Water	= 6.6 μ l

Amplification profile:	
97°C - 2 min.	1 cycle
97°C - 15 sec.	} 10 cycles
70°C - 45 sec.	
72°C - 45 sec.	
97°C - 15 sec.	} 35 cycles
64°C - 45 sec.	
72°C - 45 sec.	

Sequencing of PCR Products

The PCR products were purified using a Whatman/Polyfiltronics 100 μ l 384 well unfilter plate essentially according to the manufacturers protocol. The purified DNA was eluted in 50 μ l of distilled water. Sequencing reactions were set up using Applied Biosystems Big Dye Terminator chemistry essentially according to the manufacturers protocol. The purified PCR products were sequenced in both directions using the primer sets represented below by the nucleotide positions of their initial and final nucleotides, which correspond to positions in SEQ ID NO:1 (Figure 1). Reaction products were purified by isopropanol precipitation, and run on an Applied Biosystems 3700 DNA Analyzer.

Sequencing Primer Pairs

Fragment	Forward Primer	Reverse Primer
Fragment 1	2515 - 2536	complement of 3054 - 3035
Fragment 2	2799 - 2818	complement of 3330 - 3311
Fragment 3	2863 - 2881	complement of 3280 - 3261
Fragment 4	3071 - 3090	complement of 3487 - 3468
Fragment 5	3286 - 3305	complement of 3757 - 3738
Fragment 6	3898 - 3917	complement of 4396 - 4377
Fragment 7	4162 - 4181	complement of 4705 - 4686
Fragment 8	4435 - 4453	complement of 4942 - 4922
Fragment 9	4730 - 4750	complement of 5246 - 5227

5 Analysis of Sequences for Polymorphic Sites

Sequence information for a minimum of 80 humans was analyzed for the presence of polymorphisms using the Polyphred program (Nickerson et al., *Nucleic Acids Res.* 14:2745-2751, 1997). The presence of a polymorphism was confirmed on both strands. The polymorphisms and their locations in the FY reference genomic sequence (SEQ ID NO:1) are listed in Table 2 below.

10

Table 2. Polymorphic Sites Identified in the FY Gene

Polymorphic Site No.	PolyId(a)	Nucleotide Position	Reference Allele	Variant Allele	CDS Variant Position	AA Variant	Double PS Variant(b)
PS1	8104322	2690	C	T			
PS2	8104330	2864	G	A			
PS3	8104332	2882	A	G			
PS4	8104334	2910	C	T			
PS5	8104340	2949	C	A			
PS6	8104344	2980	G	C			
PS7	8104346	2996	C	T			
PS8	8104350	3259	T	C			
PS9(R)	8104352	3470	T	C			
PS10	8104354	3672	C	T			
PS11	8104356	3707	C	T			
PS12	8104358	3979	A	G			
PS13	8104360	3997	C	T			
PS14(R)	8104364	4140	A	G	131	D44G	
PS15	8104366	4214	C	T	205	L69F	
PS16(R)	8104370	4280	C	T	271	R91C	
PS17(R)	8104372	4313	G	A	304	A102T	
PS18	8104379	4617	C	T	608	T203M	
PS19	8104381	4618	G	A	609	T203T	T203I
PS20	8104387	4992	C	T	983	S328F	

(a) PolyId is a unique identifier assigned to each PS by Genaissance Pharmaceuticals, Inc.

(b) Double PS Variant refers to an amino acid change caused by two polymorphic sites within the same codon

5 (R) Reported previously.

EXAMPLE 2

10 This example illustrates analysis of the FY polymorphisms identified in the Index Repository for human genotypes and haplotypes.

The different genotypes containing these polymorphisms that were observed in unrelated members of the reference population are shown in Table 3 below, with the haplotype pair indicating the combination of haplotypes determined for the individual using the haplotype derivation protocol described below. In Table 3, homozygous positions are indicated by one nucleotide and heterozygous positions are indicated by two nucleotides. Missing nucleotides in any given genotype in Table 3 were
15 inferred based on linkage disequilibrium and/or Mendelian inheritance.

20

Table 3(Part 1). Genotypes and Haplotype Pairs Observed for FY Gene

Genotype Number	HAP Pair		Polymorphic Sites									
			PS1	PS2	PS3	PS4	PS5	PS6	PS7	PS8	PS9	PS10
1	10	10	C	G	A	C	C	G	C	T	T	T
5 2	21	21	C	G	A	T	C	G	T	T	T	C
3	7	7	C	G	A	C	C	G	C	T	T	C
4	14	14	C	G	A	C	C	G	C	T	T	T
5	5	5	C	G	A	C	C	G	C	T	C	T
6	12	12	C	G	A	C	C	G	C	T	T	T
10 7	10	16	C	G	A	C	C	G	C	T	T	T
8	7	1	C	G/A	A	C	C	G	C	T	T	C
9	10	14	C	G	A	C	C	G	C	T	T	T
10	21	15	C	G	A	T/C	C	G	T/C	T	T	C/T
11	5	8	C	G	A	C	C	G	C	T	C/T	T/C
15 12	5	14	C	G	A	C	C	G	C	T	C/T	T
13	10	6	C	G	A	C	C	G	C	T	T	T/C
14	18	2	C	G	A	C	C/A	G	C	T	T	T
15	15	19	C	G	A	C/T	C	G	C	T	T	T
16	21	7	C	G	A	T/C	C	G	T/C	T	T	C
20 17	5	21	C	G	A	C/T	C	G	C/T	T	C/T	T/C
18	10	13	C	G	A	C	C	G	C	T	T	T
19	10	1	C	G/A	A	C	C	G	C	T	T	T/C
20	14	1	C	G/A	A	C	C	G	C	T	T	T/C
21	10	21	C	G	A	C/T	C	G	C/T	T	T	T/C
25 22	10	18	C	G	A	C	C	G	C	T	T	T
23	18	15	C	G	A	C	C	G	C	T	T	T
24	10	7	C	G	A	C	C	G	C	T	T	T/C
25	5	23	C/T	G	A	C	C	G	C	T	C	T
26	10	5	C	G	A	C	C	G	C	T	T/C	T
30 27	10	12	C	G	A	C	C	G	C	T	T	T
28	5	22	C	G	A/G	C	C	G	C	T	C	T/C
29	5	7	C	G	A	C	C	G	C	T	C/T	T/C
30	7	22	C	G	A/G	C	C	G	C	T	T/C	C
31	21	14	C	G	A	T/C	C	G	T/C	T	T	C/T
35 32	10	3	C	G	A	C	C	G/C	C	T	T	T
33	10	9	C	G	A	C	C	G	C	T	T	T
34	10	11	C	G	A	C	C	G	C	T	T	T
35	20	2	C	G	A	T/C	C/A	G	C	T	T	T
36	10	20	C	G	A	C/T	C	G	C	T	T	T
40 37	20	7	C	G	A	T/C	C	G	C	T	T	T/C
38	10	17	C	G	A	C	C	G	C	T	T	T
39	10	4	C	G	A	C	C	G	C	T/C	T	T
40	5	18	C	G	A	C	C	G	C	T	C/T	T

45

Table 3(Part 2). Genotypes and Haplotype Pairs Observed for FY Gene

Genotype Number	HAP Pair		Polymorphic Sites									
			PS11	PS12	PS13	PS14	PS15	PS16	PS17	PS18	PS19	PS20
1	10	10	C	A	C	G	C	C	G	C	G	C
5 2	21	21	C	A	C	A	C	C	G	C	G	C
3	7	7	C	G	C	A	C	C	G	C	G	C
4	14	14	T	A	C	A	C	C	A	C	G	C
5	5	5	C	A	C	A	C	C	G	C	G	C
6	12	12	C	A	C	G	T	C	G	C	G	C
10 7	10	16	C/T	A	C	G/A	C	C/T	G/A	C	G	C
8	7	1	C	G	C	A	C	C	G	C	G	C
9	10	14	C/T	A	C	G/A	C	C	G/A	C	G	C
10	21	15	C/T	A	C	A	C	C	G	C	G	C
11	5	8	C	A/G	C/T	A	C	C	G	C	G	C
15 12	5	14	C/T	A	C	A	C	C	G/A	C	G	C
13	10	6	C	A	C	G	C	C	G	C	G	C
14	18	2	T	A	C	G/A	C	C	G/A	C	G	C
15	15	19	T/C	A	C	A	C	C	G	C	G	C
16	21	7	C	A/G	C	A	C	C	G	C	G	C
20 17	5	21	C	A	C	A	C	C	G	C	G	C
18	10	13	C	A	C	G	C/T	C	G	C	G	C/T
19	10	1	C	A/G	C	G/A	C	C	G	C	G	C
20	14	1	T/C	A/G	C	A	C	C	A/G	C	G	C
21	10	21	C	A	C	G/A	C	C	G	C	G	C
25 22	10	18	C/T	A	C	G	C	C	G	C	G	C
23	18	15	T	A	C	G/A	C	C	G	C	G	C
24	10	7	C	A/G	C	G/A	C	C	G	C	G	C
25	5	23	C	A	C	A	C	C	G	C	G	C
26	10	5	C	A	C	G/A	C	C	G	C	G	C
30 27	10	12	C	A	C	G	C/T	C	G	C	G	C
28	5	22	C	A	C	A	C	C	G	C	G	C
29	5	7	C	A/G	C	A	C	C	G	C	G	C
30	7	22	C	G/A	C	A	C	C	G	C	G	C
31	21	14	C/T	A	C	A	C	C	G/A	C	G	C
35 32	10	3	C	A	C	G	C	C	G	C	G	C
33	10	9	C	A	C	G	C	C	G	C	G/A	C
34	10	11	C	A	C	G	C	C	G	C/T	G	C
35	20	2	T	A	C	A	C	C	G/A	C	G	C
36	10	20	C/T	A	C	G/A	C	C	G	C	G	C
40 37	20	7	T/C	A/G	C	A	C	C	G	C	G	C
38	10	17	C/T	A	C	G/A	C/T	C	G	C	G	C
39	10	4	C	A	C	G	C	C	G	C	G	C
40	5	18	C/T	A	C	A/G	C	C	G	C	G	C

The haplotype pairs shown in Table 3 were estimated from the unphased genotypes using a computer-implemented extension of Clark's algorithm (Clark, A.G. 1990 *Mol Bio Evol* 7, 111-122) for assigning haplotypes to unrelated individuals in a population sample, as described in

PCT/US01/12831, filed April 18, 2001. In this method, haplotypes are assigned directly from individuals who are homozygous at all sites or heterozygous at no more than one of the variable sites.

This list of haplotypes is then used to deconvolute the unphased genotypes in the remaining (multiply heterozygous) individuals. In the present analysis, the list of haplotypes was augmented with haplotypes obtained from two families (one three-generation Caucasian family and one two-generation

African-American family).

By following this protocol, it was determined that the Index Repository examined herein and, by extension, the general population contains the 23 human FY haplotypes shown in Table 4 below.

An FY isogene defined by a full-haplotype shown in Table 4 below comprises the regions of the SEQ ID NOS indicated in Table 4, with their corresponding set of polymorphic locations and identities, which are also set forth in Table 4.

Table 4 (Part 1). Haplotypes of the FY gene.

Regions	PS No.(b)	PS Position(c)	Haplotype Number(d)									
			1	2	3	4	5	6	7	8	9	10
10	Examined(a)											
	2486-3779	1	2690/30	C	C	C	C	C	C	C	C	C
	2486-3779	2	2864/150	A	G	G	G	G	G	G	G	G
	2486-3779	3	2882/270	A	A	A	A	A	A	A	A	A
	2486-3779	4	2910/390	C	C	C	C	C	C	C	C	C
15	2486-3779	5	2949/510	C	A	C	C	C	C	C	C	C
	2486-3779	6	2980/630	G	G	C	G	G	G	G	G	G
	2486-3779	7	2996/750	C	C	C	C	C	C	C	C	C
	2486-3779	8	3259/870	T	T	T	C	T	T	T	T	T
	2486-3779	9	3470/990	T	T	T	T	C	T	T	T	T
20	2486-3779	10	3672/1110	C	T	T	T	T	C	C	C	T
	2486-3779	11	3707/1230	C	T	C	C	C	C	C	C	C
	3872-5302	12	3979/1350	G	A	A	A	A	A	G	G	A
	3872-5302	13	3997/1470	C	C	C	C	C	C	C	T	C
	3872-5302	14	4140/1590	A	A	G	G	A	G	A	A	G
25	3872-5302	15	4214/1710	C	C	C	C	C	C	C	C	C
	3872-5302	16	4280/1830	C	C	C	C	C	C	C	C	C
	3872-5302	17	4313/1950	G	A	G	G	G	G	G	G	G
	3872-5302	18	4617/2070	C	C	C	C	C	C	C	C	C
	3872-5302	19	4618/2190	G	G	G	G	G	G	G	G	A
30	3872-5302	20	4992/2310	C	C	C	C	C	C	C	C	C

Table 4 (Part 2). Haplotypes of the FY gene.

Regions	PS	PS	Haplotype Number(d)									
Examined(a)	No.(b)	Position(c)	11	12	13	14	15	16	17	18	19	20
5	2486-3779	1	2690/30	C	C	C	C	C	C	C	C	C
	2486-3779	2	2864/150	G	G	G	G	G	G	G	G	G
	2486-3779	3	2882/270	A	A	A	A	A	A	A	A	A
	2486-3779	4	2910/390	C	C	C	C	C	C	C	T	T
	2486-3779	5	2949/510	C	C	C	C	C	C	C	C	C
	2486-3779	6	2980/630	G	G	G	G	G	G	G	G	G
10	2486-3779	7	2996/750	C	C	C	C	C	C	C	C	C
	2486-3779	8	3259/870	T	T	T	T	T	T	T	T	T
	2486-3779	9	3470/990	T	T	T	T	T	T	T	T	T
	2486-3779	10	3672/1110	T	T	T	T	T	T	T	T	T
	2486-3779	11	3707/1230	C	C	C	T	T	T	T	C	T
15	3872-5302	12	3979/1350	A	A	A	A	A	A	A	A	A
	3872-5302	13	3997/1470	C	C	C	C	C	C	C	C	C
	3872-5302	14	4140/1590	G	G	G	A	A	A	A	G	A
	3872-5302	15	4214/1710	C	T	T	C	C	C	T	C	C
	3872-5302	16	4280/1830	C	C	C	C	C	T	C	C	C
20	3872-5302	17	4313/1950	G	G	G	A	G	A	G	G	G
	3872-5302	18	4617/2070	T	C	C	C	C	C	C	C	C
	3872-5302	19	4618/2190	G	G	G	G	G	G	G	G	G
	3872-5302	20	4992/2310	C	C	T	C	C	C	C	C	C

25 Table 4 (Part 3). Haplotypes of the FY gene.

Regions	PS	PS	Haplotype Number(d)		
Examined(a)	No.(b)	Position(c)	21	22	23
30	2486-3779	1	2690/30	C	T
	2486-3779	2	2864/150	G	G
	2486-3779	3	2882/270	A	A
	2486-3779	4	2910/390	T	C
	2486-3779	5	2949/510	C	C
	2486-3779	6	2980/630	G	G
	2486-3779	7	2996/750	T	C
35	2486-3779	8	3259/870	T	T
	2486-3779	9	3470/990	T	C
	2486-3779	10	3672/1110	C	T
	2486-3779	11	3707/1230	C	C
	3872-5302	12	3979/1350	A	A
40	3872-5302	13	3997/1470	C	C
	3872-5302	14	4140/1590	A	A
	3872-5302	15	4214/1710	C	C
	3872-5302	16	4280/1830	C	C
	3872-5302	17	4313/1950	G	G
45	3872-5302	18	4617/2070	C	C
	3872-5302	19	4618/2190	G	G
	3872-5302	20	4992/2310	C	C

(a) Region examined represents the nucleotide positions defining the start and stop positions within SEQ ID NO:1 of the regions sequenced;

(b) PS = polymorphic site;

(c) Position of PS within the indicated SEQ ID NO, with the 1st position number referring to SEQ ID NO:1 and the 2nd position number referring to SEQ ID NO:84, a modified version of SEQ ID NO:1 that comprises the context sequence of each polymorphic site, PS1-PS20, to facilitate electronic searching of the haplotypes;

(d) Alleles for FY haplotypes are presented 5' to 3' in each column.

SEQ ID NO:1 refers to Figure 1, with the two alternative allelic variants of each polymorphic site indicated by the appropriate nucleotide symbol. SEQ ID NO:84 is a modified version of SEQ ID NO:1 that shows the context sequence of each of PS1-PS20 in a uniform format to facilitate electronic searching of the FY haplotypes. For each polymorphic site, SEQ ID NO:84 contains a block of 60 bases of the nucleotide sequence encompassing the centrally-located polymorphic site at the 30th position, followed by 60 bases of unspecified sequence to represent that each polymorphic site is separated by genomic sequence whose composition is defined elsewhere herein.

Table 5 below shows the percent of chromosomes characterized by a given FY haplotype for all unrelated individuals in the Index Repository for which haplotype data was obtained. The percent of these unrelated individuals who have a given FY haplotype pair is shown in Table 6. In Tables 5 and 6, the "Total" column shows this frequency data for all of these unrelated individuals, while the other columns show the frequency data for these unrelated individuals categorized according to their self-identified ethnogeographic origin. Abbreviations used in Tables 5 and 6 are AF = African Descent, AS = Asian, CA = Caucasian, HL = Hispanic-Latino, and AM = Native American.

Table 5 Frequency of Observed FY Haplotypes In Unrelated Individuals

HAP No.	HAP ID	Total	CA	AF	AS	HL	AM
1	8107557	1.83	0.0	0.0	0.0	8.33	0.0
2	8107559	1.22	4.76	0.0	0.0	0.0	0.0
3	8107568	0.61	0.0	0.0	2.5	0.0	0.0
4	8107564	0.61	0.0	0.0	0.0	2.78	0.0
5	8107548	15.85	0.0	62.5	0.0	2.78	0.0
6	8107565	0.61	0.0	0.0	2.5	0.0	0.0
7	8107551	6.1	9.52	5.0	0.0	5.56	33.33
8	8107563	0.61	0.0	2.5	0.0	0.0	0.0
9	8107566	0.61	0.0	0.0	2.5	0.0	0.0
10	8107547	36.59	42.86	2.5	70.0	30.56	33.33
11	8107567	0.61	0.0	0.0	2.5	0.0	0.0
12	8107555	3.05	0.0	0.0	12.5	0.0	0.0
13	8107561	0.61	0.0	0.0	2.5	0.0	0.0
14	8107552	4.88	2.38	2.5	0.0	16.67	0.0
15	8107556	1.83	4.76	0.0	0.0	2.78	0.0
16	8107562	0.61	0.0	0.0	0.0	2.78	0.0
17	8107569	0.61	0.0	0.0	2.5	0.0	0.0
18	8107553	3.66	7.14	5.0	0.0	2.78	0.0
19	8107560	0.61	2.38	0.0	0.0	0.0	0.0
20	8107549	7.32	16.67	0.0	2.5	5.56	33.33
21	8107550	6.71	9.52	0.0	0.0	19.44	0.0
22	8107554	3.05	0.0	12.5	0.0	0.0	0.0
23	8107558	1.83	0.0	7.5	0.0	0.0	0.0

Table 6. Frequency of Observed FY Haplotype Pairs In Unrelated Individuals

	HAP1	HAP2	Total	CA	AF	AS	HL	AM
5	10	10	17.07	19.05	0.0	45.0	5.56	0.0
	21	21	1.22	0.0	0.0	0.0	5.56	0.0
	7	7	1.22	0.0	0.0	0.0	0.0	33.33
	14	14	2.44	0.0	0.0	0.0	11.11	0.0
10	5	5	7.32	0.0	30.0	0.0	0.0	0.0
	12	12	1.22	0.0	0.0	5.0	0.0	0.0
	10	16	1.22	0.0	0.0	0.0	5.56	0.0
	7	1	1.22	0.0	0.0	0.0	5.56	0.0
15	10	14	1.22	0.0	0.0	0.0	5.56	0.0
	21	15	1.22	0.0	0.0	0.0	5.56	0.0
	5	8	1.22	0.0	5.0	0.0	0.0	0.0
	5	14	1.22	0.0	5.0	0.0	0.0	0.0
20	10	6	1.22	0.0	0.0	5.0	0.0	0.0
	18	2	1.22	4.76	0.0	0.0	0.0	0.0
	15	19	1.22	4.76	0.0	0.0	0.0	0.0
	21	7	2.44	4.76	0.0	0.0	5.56	0.0
25	5	21	1.22	0.0	0.0	0.0	5.56	0.0
	10	13	1.22	0.0	0.0	5.0	0.0	0.0
	10	1	1.22	0.0	0.0	0.0	5.56	0.0
	14	1	1.22	0.0	0.0	0.0	5.56	0.0
30	10	21	4.88	9.52	0.0	0.0	11.11	0.0
	10	18	2.44	4.76	0.0	0.0	5.56	0.0
	18	15	1.22	4.76	0.0	0.0	0.0	0.0
	10	7	2.44	9.52	0.0	0.0	0.0	0.0
35	5	23	3.66	0.0	15.0	0.0	0.0	0.0
	10	5	1.22	0.0	5.0	0.0	0.0	0.0
	10	12	3.66	0.0	0.0	15.0	0.0	0.0
	5	22	4.88	0.0	20.0	0.0	0.0	0.0
40	5	7	1.22	0.0	5.0	0.0	0.0	0.0
	7	22	1.22	0.0	5.0	0.0	0.0	0.0
	21	14	1.22	4.76	0.0	0.0	0.0	0.0
	10	3	1.22	0.0	0.0	5.0	0.0	0.0
45	10	9	1.22	0.0	0.0	5.0	0.0	0.0
	10	11	1.22	0.0	0.0	5.0	0.0	0.0
	20	2	1.22	4.76	0.0	0.0	0.0	0.0
	10	20	12.2	23.81	0.0	5.0	11.11	66.67
50	20	7	1.22	4.76	0.0	0.0	0.0	0.0
	10	17	1.22	0.0	0.0	5.0	0.0	0.0
	10	4	1.22	0.0	0.0	0.0	5.56	0.0
	5	18	2.44	0.0	10.0	0.0	0.0	0.0

The size and composition of the Index Repository were chosen to represent the genetic diversity across and within four major population groups comprising the general United States population. For example, as described in Table 1 above, this repository contains approximately equal sample sizes of African-descent, Asian-American, European-American, and Hispanic-Latino population groups. Almost all individuals representing each group had all four grandparents with the same ethnogeographic background. The number of unrelated individuals in the Index Repository provides a sample size that is sufficient to detect SNPs and haplotypes that occur in the general population with high statistical certainty. For instance, a haplotype that occurs with a frequency of 5%

in the general population has a probability higher than 99.9% of being observed in a sample of 80 individuals from the general population. Similarly, a haplotype that occurs with a frequency of 10% in a specific population group has a 99% probability of being observed in a sample of 20 individuals from that population group. In addition, the size and composition of the Index Repository means that the relative frequencies determined therein for the haplotypes and haplotype pairs of the FY gene are likely to be similar to the relative frequencies of these FY haplotypes and haplotype pairs in the general U.S. population and in the four population groups represented in the Index Repository. The genetic diversity observed for the three Native Americans is presented because it is of scientific interest, but due to the small sample size it lacks statistical significance.

In view of the above, it will be seen that the several advantages of the invention are achieved and other advantageous results attained.

As various changes could be made in the above methods and compositions without departing from the scope of the invention, it is intended that all matter contained in the above description and shown in the accompanying drawings shall be interpreted as illustrative and not in a limiting sense.

All references cited in this specification, including patents and patent applications, are hereby incorporated in their entirety by reference. The discussion of references herein is intended merely to summarize the assertions made by their authors and no admission is made that any reference constitutes prior art. Applicants reserve the right to challenge the accuracy and pertinency of the cited references.

What is Claimed is:

1. A method for haplotyping the Duffy blood group (FY) gene of an individual, which comprises determining which of the FY haplotypes shown in the table immediately below defines one copy of the individual's FY gene, wherein the determining step comprises identifying the phased sequence of nucleotides present at each of PS1-PS20 on at least one copy of the individual's FY gene, and wherein each of the FY haplotypes comprises a sequence of polymorphisms whose positions and identities are set forth in the table immediately below:

PS		Haplotype Number(c) (Part 1)										
10	No.(a)	PS Position(b)	1	2	3	4	5	6	7	8	9	10
	1	2690	C	C	C	C	C	C	C	C	C	C
	2	2864	A	G	G	G	G	G	G	G	G	G
	3	2882	A	A	A	A	A	A	A	A	A	A
	4	2910	C	C	C	C	C	C	C	C	C	C
15	5	2949	C	A	C	C	C	C	C	C	C	C
	6	2980	G	G	C	G	G	G	G	G	G	G
	7	2996	C	C	C	C	C	C	C	C	C	C
	8	3259	T	T	T	C	T	T	T	T	T	T
	9	3470	T	T	T	T	C	T	T	T	T	T
20	10	3672	C	T	T	T	T	C	C	C	T	T
	11	3707	C	T	C	C	C	C	C	C	C	C
	12	3979	G	A	A	A	A	A	G	G	A	A
	13	3997	C	C	C	C	C	C	C	T	C	C
	14	4140	A	A	G	G	A	G	A	A	G	G
25	15	4214	C	C	C	C	C	C	C	C	C	C
	16	4280	C	C	C	C	C	C	C	C	C	C
	17	4313	G	A	G	G	G	G	G	G	G	G
	18	4617	C	C	C	C	C	C	C	C	C	C
	19	4618	G	G	G	G	G	G	G	G	A	G
30	20	4992	C	C	C	C	C	C	C	C	C	C

PS		Haplotype Number(c) (Part 2)									
No.(a)	Position(b)	11	12	13	14	15	16	17	18	19	20
1	2690	C	C	C	C	C	C	C	C	C	C
2	2864	G	G	G	G	G	G	G	G	G	G
5 3	2882	A	A	A	A	A	A	A	A	A	A
4	2910	C	C	C	C	C	C	C	C	T	T
5	2949	C	C	C	C	C	C	C	C	C	C
6	2980	G	G	G	G	G	G	G	G	G	G
7	2996	C	C	C	C	C	C	C	C	C	C
10 8	3259	T	T	T	T	T	T	T	T	T	T
9	3470	T	T	T	T	T	T	T	T	T	T
10	3672	T	T	T	T	T	T	T	T	T	T
11	3707	C	C	C	T	T	T	T	T	C	T
12	3979	A	A	A	A	A	A	A	A	A	A
15 13	3997	C	C	C	C	C	C	C	C	C	C
14	4140	G	G	G	A	A	A	A	G	A	A
15	4214	C	T	T	C	C	C	T	C	C	C
16	4280	C	C	C	C	C	T	C	C	C	C
17	4313	G	G	G	A	G	A	G	G	G	G
20 18	4617	T	C	C	C	C	C	C	C	C	C
19	4618	G	G	G	G	G	G	G	G	G	G
20	4992	C	C	T	C	C	C	C	C	C	C

PS		Haplotype Number(c) (Part 3)		
No.(a)	Position(b)	21	22	23
1	2690	C	C	T
2	2864	G	G	G
3	2882	A	G	A
4	2910	T	C	C
30 5	2949	C	C	C
6	2980	G	G	G
7	2996	T	C	C
8	3259	T	T	T
9	3470	T	C	C
35 10	3672	C	C	T
11	3707	C	C	C
12	3979	A	A	A
13	3997	C	C	C
14	4140	A	A	A
40 15	4214	C	C	C
16	4280	C	C	C
17	4313	G	G	G
18	4617	C	C	C
19	4618	G	G	G
45 20	4992	C	C	C

(a) PS = polymorphic site;

(b) Position of PS within SEQ ID NO:1;

(c) Alleles for haplotypes are presented 5' to 3' in each column.

50

2. A method for haplotyping the Duffy blood group (FY) gene of an individual, which comprises determining which of the FY haplotype pairs shown in the table immediately below defines both copies of the individual's FY gene, wherein the determining step comprises identifying

the phased sequence of nucleotides present at each of PS1-PS20 on both copies of the individual's FY gene, and wherein each of the FY haplotype pairs consists of first and second haplotypes which comprise first and second sequences of polymorphisms whose positions and identities are set forth in the table immediately below:

PS No.(a)	PS Position(b)	Haplotype Pair(c) (Part 1)							
		10/10	21/21	7/7	14/14	5/5	12/12	10/16	7/1
1	2690	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
2	2864	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A
3	2882	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
4	2910	C/C	T/T	C/C	C/C	C/C	C/C	C/C	C/C
5	2949	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
6	2980	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
7	2996	C/C	T/T	C/C	C/C	C/C	C/C	C/C	C/C
8	3259	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
9	3470	T/T	T/T	T/T	T/T	C/C	T/T	T/T	T/T
10	3672	T/T	C/C	C/C	T/T	T/T	T/T	T/T	C/C
11	3707	C/C	C/C	C/C	T/T	C/C	C/C	C/T	C/C
12	3979	A/A	A/A	G/G	A/A	A/A	A/A	A/A	G/G
13	3997	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
14	4140	G/G	A/A	A/A	A/A	A/A	G/G	G/A	A/A
15	4214	C/C	C/C	C/C	C/C	C/C	T/T	C/C	C/C
16	4280	C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/C
17	4313	G/G	G/G	G/G	A/A	G/G	G/G	G/A	G/G
18	4617	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
19	4618	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
20	4992	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C

PS		Haplotype Pair(c) (Part 2)								
No.(a)	Position(b)	10/14	21/15	5/8	5/14	10/6	18/2	15/19	21/7	
5	1	2690	C/C	C/C	C/C	C/C	C/C	C/C	C/C	
	2	2864	G/G	G/G	G/G	G/G	G/G	G/G	G/G	
	3	2882	A/A	A/A	A/A	A/A	A/A	A/A	A/A	
	4	2910	C/C	T/C	C/C	C/C	C/C	C/C	T/C	
	5	2949	C/C	C/C	C/C	C/C	C/C	C/A	C/C	
10	6	2980	G/G	G/G	G/G	G/G	G/G	G/G	G/G	
	7	2996	C/C	T/C	C/C	C/C	C/C	C/C	T/C	
	8	3259	T/T	T/T	T/T	T/T	T/T	T/T	T/T	
	9	3470	T/T	T/T	C/T	C/T	T/T	T/T	T/T	
	10	3672	T/T	C/T	T/C	T/T	T/C	T/T	C/C	
15	11	3707	C/T	C/T	C/C	C/T	C/C	T/T	C/C	
	12	3979	A/A	A/A	A/G	A/A	A/A	A/A	A/G	
	13	3997	C/C	C/C	C/T	C/C	C/C	C/C	C/C	
	14	4140	G/A	A/A	A/A	A/A	G/G	G/A	A/A	
	15	4214	C/C	C/C	C/C	C/C	C/C	C/C	C/C	
20	16	4280	C/C	C/C	C/C	C/C	C/C	C/C	C/C	
	17	4313	G/A	G/G	G/G	G/A	G/G	G/A	G/G	
	18	4617	C/C	C/C	C/C	C/C	C/C	C/C	C/C	
	19	4618	G/G	G/G	G/G	G/G	G/G	G/G	G/G	
	20	4992	C/C	C/C	C/C	C/C	C/C	C/C	C/C	

PS		Haplotype Pair(c) (Part 3)								
No.(a)	Position(b)	5/21	10/13	10/1	14/1	10/21	10/18	18/15	10/7	
25	1	2690	C/C	C/C	C/C	C/C	C/C	C/C	C/C	
	2	2864	G/G	G/G	G/A	G/A	G/G	G/G	G/G	
	3	2882	A/A	A/A	A/A	A/A	A/A	A/A	A/A	
	4	2910	C/T	C/C	C/C	C/C	C/T	C/C	C/C	
	5	2949	C/C	C/C	C/C	C/C	C/C	C/C	C/C	
30	6	2980	G/G	G/G	G/G	G/G	G/G	G/G	G/G	
	7	2996	C/T	C/C	C/C	C/C	C/T	C/C	C/C	
	8	3259	T/T	T/T	T/T	T/T	T/T	T/T	T/T	
	9	3470	C/T	T/T	T/T	T/T	T/T	T/T	T/T	
	10	3672	T/C	T/T	T/C	T/C	T/T	T/T	T/C	
35	11	3707	C/C	C/C	C/C	T/C	C/C	T/T	C/C	
	12	3979	A/A	A/A	A/G	A/G	A/A	A/A	A/G	
	13	3997	C/C	C/C	C/C	C/C	C/C	C/C	C/C	
	14	4140	A/A	G/G	G/A	A/A	G/A	G/G	G/A	
	15	4214	C/C	C/T	C/C	C/C	C/C	C/C	C/C	
40	16	4280	C/C	C/C	C/C	C/C	C/C	C/C	C/C	
	17	4313	G/G	G/G	G/G	A/G	G/G	G/G	G/G	
	18	4617	C/C	C/C	C/C	C/C	C/C	C/C	C/C	
	19	4618	G/G	G/G	G/G	G/G	G/G	G/G	G/G	
	20	4992	C/C	C/T	C/C	C/C	C/C	C/C	C/C	

PS		Haplotype Pair(c) (Part 4)							
No.(a)	Position(b)	5/23	10/5	10/12	5/22	5/7	7/22	21/14	10/3
5	1	2690	C/T	C/C	C/C	C/C	C/C	C/C	C/C
	2	2864	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	3	2882	A/A	A/A	A/A	A/G	A/A	A/G	A/A
	4	2910	C/C	C/C	C/C	C/C	C/C	T/C	C/C
	5	2949	C/C	C/C	C/C	C/C	C/C	C/C	C/C
10	6	2980	G/G	G/G	G/G	G/G	G/G	G/G	G/C
	7	2996	C/C	C/C	C/C	C/C	C/C	T/C	C/C
	8	3259	T/T	T/T	T/T	T/T	T/T	T/T	T/T
	9	3470	C/C	T/C	T/T	C/C	C/T	T/C	T/T
	10	3672	T/T	T/T	T/T	T/C	T/C	C/T	T/T
15	11	3707	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	12	3979	A/A	A/A	A/A	A/A	A/G	G/A	A/A
	13	3997	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	14	4140	A/A	G/A	G/G	A/A	A/A	A/A	G/G
	15	4214	C/C	C/C	C/T	C/C	C/C	C/C	C/C
20	16	4280	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	17	4313	G/G	G/G	G/G	G/G	G/G	G/A	G/G
	18	4617	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	19	4618	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	20	4992	C/C	C/C	C/C	C/C	C/C	C/C	C/C

PS		Haplotype Pair(c) (Part 5)							
No.(a)	Position(b)	10/9	10/11	20/2	10/20	20/7	10/17	10/4	5/18
25	1	2690	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	2	2864	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	3	2882	A/A	A/A	A/A	A/A	A/A	A/A	A/A
	4	2910	C/C	C/C	T/C	C/T	T/C	C/C	C/C
	5	2949	C/C	C/C	C/A	C/C	C/C	C/C	C/C
30	6	2980	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	7	2996	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	8	3259	T/T	T/T	T/T	T/T	T/T	T/C	T/T
	9	3470	T/T	T/T	T/T	T/T	T/T	T/T	C/T
	10	3672	T/T	T/T	T/T	T/T	T/C	T/T	T/T
35	11	3707	C/C	C/C	T/T	C/T	T/C	C/C	C/T
	12	3979	A/A	A/A	A/A	A/A	A/G	A/A	A/A
	13	3997	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	14	4140	G/G	G/G	A/A	G/A	A/A	G/A	A/G
	15	4214	C/C	C/C	C/C	C/C	C/C	C/T	C/C
40	16	4280	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	17	4313	G/G	G/G	G/A	G/G	G/G	G/G	G/G
	18	4617	C/C	C/T	C/C	C/C	C/C	C/C	C/C
	19	4618	G/A	G/G	G/G	G/G	G/G	G/G	G/G
	20	4992	C/C	C/C	C/C	C/C	C/C	C/C	C/C

(a) PS = polymorphic site;

(b) Position of PS in SEQ ID NO:1;

(c) Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5' to 3' as 1st polymorphism/2nd polymorphism in each column.

3. A method for genotyping the Duffy blood group (FY) gene of an individual, comprising determining for the two copies of the FY gene present in the individual the identity of the nucleotide pair at one or more polymorphic sites (PS) selected from the group consisting of

- 5 PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS10, PS11, PS12, PS13, PS15, PS18, PS19 and
PS20, wherein the one or more polymorphic sites (PS) have the position and alternative alleles
shown in SEQ ID NO:1.
4. The method of claim 3, wherein the determining step comprises:
- (a) isolating from the individual a nucleic acid mixture comprising both copies of the FY
gene, or a fragment thereof, that are present in the individual;
 - (b) amplifying from the nucleic acid mixture a target region containing one of the selected
5 polymorphic sites;
 - (c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region,
wherein the oligonucleotide is designed for genotyping the selected polymorphic site in
the target region;
 - (d) performing a nucleic acid template-dependent, primer extension reaction on the
10 hybridized oligonucleotide in the presence of at least one terminator of the reaction,
wherein the terminator is complementary to one of the alternative nucleotides present at
the selected polymorphic site; and
 - (e) detecting the presence and identity of the terminator in the extended oligonucleotide.
5. The method of claim 3, which comprises determining for the two copies of the FY gene present
in the individual the identity of the nucleotide pair at each of PS1-PS20.
6. A method for haplotyping the Duffy blood group (FY) gene of an individual which comprises
determining, for one copy of the FY gene present in the individual, the identity of the nucleotide
at two or more polymorphic sites (PS) selected from the group consisting of PS1, PS2, PS3,
PS4, PS5, PS6, PS7, PS8, PS10, PS11, PS12, PS13, PS15, PS18, PS19 and PS20, wherein the
selected PS have the position and alternative alleles shown in SEQ ID NO:1.
7. The method of claim 6, further comprising determining the identity of the nucleotide at one or
more polymorphic sites selected from the group consisting of PS9, PS14, PS16 and PS17,
wherein the one or more polymorphic sites (PS) have the position and alternative alleles shown
in SEQ ID NO:1.
8. The method of claim 6, wherein the determining step comprises:
- (a) isolating from the individual a nucleic acid sample containing only one of the two copies
of the FY gene, or a fragment thereof, that is present in the individual;
 - (b) amplifying from the nucleic acid sample a target region containing one of the selected
5 polymorphic sites;
 - (c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region,
wherein the oligonucleotide is designed for haplotyping the selected polymorphic site in
the target region;
 - (d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized
10 oligonucleotide in the presence of at least one terminator of the reaction, wherein the

terminator is complementary to one of the alternative nucleotides present at the selected polymorphic site; and

- (e) detecting the presence and identity of the terminator in the extended oligonucleotide.
9. A method for predicting a haplotype pair for the Duffy blood group (FY) gene of an individual comprising:
- (a) identifying a FY genotype for the individual, wherein the genotype comprises the nucleotide pair at two or more polymorphic sites (PS) selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS10, PS11, PS12, PS13, PS15, PS18, PS19 and PS20, wherein the selected PS have the position and alternative alleles shown in SEQ ID NO:1;
- (b) comparing the genotype to the haplotype pair data set forth in the table immediately below; and
- (c) determining which haplotype pair is consistent with the genotype of the individual and with the haplotype pair data

PS		Haplotype Pair(c) (Part 1)								
	No.(a)	Position(b)	10/10	21/21	7/7	14/14	5/5	12/12	10/16	7/1
15	1	2690	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	2	2864	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A
	3	2882	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
	4	2910	C/C	T/T	C/C	C/C	C/C	C/C	C/C	C/C
	5	2949	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
20	6	2980	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	7	2996	C/C	T/T	C/C	C/C	C/C	C/C	C/C	C/C
	8	3259	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
	9	3470	T/T	T/T	T/T	T/T	C/C	T/T	T/T	T/T
	10	3672	T/T	C/C	C/C	T/T	T/T	T/T	T/T	C/C
25	11	3707	C/C	C/C	C/C	T/T	C/C	C/C	C/T	C/C
	12	3979	A/A	A/A	G/G	A/A	A/A	A/A	A/A	G/G
	13	3997	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	14	4140	G/G	A/A	A/A	A/A	A/A	G/G	G/A	A/A
	15	4214	C/C	C/C	C/C	C/C	C/C	T/T	C/C	C/C
30	16	4280	C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/C
	17	4313	G/G	G/G	G/G	A/A	G/G	G/G	G/A	G/G
	18	4617	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	19	4618	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	20	4992	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
35										

	PS		Haplotype Pair(c) (Part 2)							
	No.(a)	Position(b)	10/14	21/15	5/8	5/14	10/6	18/2	15/19	21/7
40	1	2690	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	2	2864	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	3	2882	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
	4	2910	C/C	T/C	C/C	C/C	C/C	C/C	C/T	T/C
	5	2949	C/C	C/C	C/C	C/C	C/C	C/A	C/C	C/C
45	6	2980	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	7	2996	C/C	T/C	C/C	C/C	C/C	C/C	C/C	T/C
	8	3259	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
	9	3470	T/T	T/T	C/T	C/T	T/T	T/T	T/T	T/T
	10	3672	T/T	C/T	T/C	T/T	T/C	T/T	T/T	C/C
50	11	3707	C/T	C/T	C/C	C/T	C/C	T/T	T/C	C/C
	12	3979	A/A	A/A	A/G	A/A	A/A	A/A	A/A	A/G
	13	3997	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/C
	14	4140	G/A	A/A	A/A	A/A	G/G	G/A	A/A	A/A
	15	4214	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
55	16	4280	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	17	4313	G/A	G/G	G/G	G/A	G/G	G/A	G/G	G/G
	18	4617	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	19	4618	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	20	4992	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C

	PS		Haplotype Pair(c) (Part 3)							
	No.(a)	Position(b)	5/21	10/13	10/1	14/1	10/21	10/18	18/15	10/7
60	1	2690	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	2	2864	G/G	G/G	G/A	G/A	G/G	G/G	G/G	G/G
	3	2882	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
	4	2910	C/T	C/C	C/C	C/C	C/T	C/C	C/C	C/C
	5	2949	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
65	6	2980	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	7	2996	C/T	C/C	C/C	C/C	C/T	C/C	C/C	C/C
	8	3259	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
	9	3470	C/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
	10	3672	T/C	T/T	T/C	T/C	T/C	T/T	T/T	T/C
70	11	3707	C/C	C/C	C/C	T/C	C/C	C/T	T/T	C/C
	12	3979	A/A	A/A	A/G	A/G	A/A	A/A	A/A	A/G
	13	3997	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	14	4140	A/A	G/G	G/A	A/A	G/A	G/G	G/A	G/A
	15	4214	C/C	C/T	C/C	C/C	C/C	C/C	C/C	C/C
75	16	4280	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	17	4313	G/G	G/G	G/G	A/G	G/G	G/G	G/G	G/G
	18	4617	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	19	4618	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	20	4992	C/C	C/T	C/C	C/C	C/C	C/C	C/C	C/C

		Haplotype Pair(c) (Part 4)								
PS	PS									
No.(a)	Position(b)	5/23	10/5	10/12	5/22	5/7	7/22	21/14	10/3	
85	1	2690	C/T	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	2	2864	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	3	2882	A/A	A/A	A/A	A/G	A/A	A/G	A/A	A/A
	4	2910	C/C	C/C	C/C	C/C	C/C	C/C	T/C	C/C
90	5	2949	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	6	2980	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/C
	7	2996	C/C	C/C	C/C	C/C	C/C	C/C	T/C	C/C
	8	3259	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
	9	3470	C/C	T/C	T/T	C/C	C/T	T/C	T/T	T/T
95	10	3672	T/T	T/T	T/T	T/C	T/C	C/C	C/T	T/T
	11	3707	C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/C
	12	3979	A/A	A/A	A/A	A/A	A/G	G/A	A/A	A/A
	13	3997	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	14	4140	A/A	G/A	G/G	A/A	A/A	A/A	A/A	G/G
100	15	4214	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/C
	16	4280	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	17	4313	G/G	G/G	G/G	G/G	G/G	G/G	G/A	G/G
	18	4617	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	19	4618	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
105	20	4992	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C

		Haplotype Pair(c) (Part 5)								
PS	PS									
No.(a)	Position(b)	10/9	10/11	20/2	10/20	20/7	10/17	10/4	5/18	
110	1	2690	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	2	2864	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	3	2882	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
	4	2910	C/C	C/C	T/C	C/T	T/C	C/C	C/C	C/C
	5	2949	C/C	C/C	C/A	C/C	C/C	C/C	C/C	C/C
	6	2980	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
115	7	2996	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	8	3259	T/T	T/T	T/T	T/T	T/T	T/T	T/C	T/T
	9	3470	T/T	T/T	T/T	T/T	T/T	T/T	T/T	C/T
	10	3672	T/T	T/T	T/T	T/T	T/C	T/T	T/T	T/T
	11	3707	C/C	C/C	T/T	C/T	T/C	C/T	C/C	C/T
120	12	3979	A/A	A/A	A/A	A/A	A/G	A/A	A/A	A/A
	13	3997	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	14	4140	G/G	G/G	A/A	G/A	A/A	G/A	G/G	A/G
	15	4214	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C
	16	4280	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
125	17	4313	G/G	G/G	G/A	G/G	G/G	G/G	G/G	G/G
	18	4617	C/C	C/T	C/C	C/C	C/C	C/C	C/C	C/C
	19	4618	G/A	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	20	4992	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C

130

(a) PS = polymorphic site;

(b) Position of PS in SEQ ID NO:1;

(c) Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5' to 3' as 1st polymorphism/2nd polymorphism in each column.

10. The method of claim 9, wherein the identified genotype of the individual comprises the nucleotide pair at each of PS1-PS20, which have the position and alternative alleles shown in SEQ ID NO:1.

11. A method for identifying an association between a trait and at least one haplotype or haplotype pair of the Duffy blood group (FY) gene which comprises comparing the frequency of the haplotype or haplotype pair in a population exhibiting the trait with the frequency of the haplotype or haplotype pair in a reference population, wherein the haplotype is selected from haplotypes 1-23 shown in the table presented immediately below, wherein each of the haplotypes comprises a sequence of polymorphisms whose positions and identities are set forth in the table immediately below:

	PS		Haplotype Number(c) (Part 1)									
	No.(a)	Position(b)	1	2	3	4	5	6	7	8	9	10
15	1	2690	C	C	C	C	C	C	C	C	C	C
	2	2864	A	G	G	G	G	G	G	G	G	G
	3	2882	A	A	A	A	A	A	A	A	A	A
	4	2910	C	C	C	C	C	C	C	C	C	C
	5	2949	C	A	C	C	C	C	C	C	C	C
	6	2980	G	G	C	G	G	G	G	G	G	G
20	7	2996	C	C	C	C	C	C	C	C	C	C
	8	3259	T	T	T	C	T	T	T	T	T	T
	9	3470	T	T	T	T	C	T	T	T	T	T
	10	3672	C	T	T	T	T	C	C	C	T	T
	11	3707	C	T	C	C	C	C	C	C	C	C
25	12	3979	G	A	A	A	A	A	G	G	A	A
	13	3997	C	C	C	C	C	C	C	T	C	C
	14	4140	A	A	G	G	A	G	A	A	G	G
	15	4214	C	C	C	C	C	C	C	C	C	C
	16	4280	C	C	C	C	C	C	C	C	C	C
30	17	4313	G	A	G	G	G	G	G	G	G	G
	18	4617	C	C	C	C	C	C	C	C	C	C
	19	4618	G	G	G	G	G	G	G	G	A	G
	20	4992	C	C	C	C	C	C	C	C	C	C

PS		Haplotype Number(c) (Part 2)									
No.(a)	Position(b)	11	12	13	14	15	16	17	18	19	20
1	2690	C	C	C	C	C	C	C	C	C	C
2	2864	G	G	G	G	G	G	G	G	G	G
3	2882	A	A	A	A	A	A	A	A	A	A
4	2910	C	C	C	C	C	C	C	C	T	T
5	2949	C	C	C	C	C	C	C	C	C	C
6	2980	G	G	G	G	G	G	G	G	G	G
7	2996	C	C	C	C	C	C	C	C	C	C
8	3259	T	T	T	T	T	T	T	T	T	T
9	3470	T	T	T	T	T	T	T	T	T	T
10	3672	T	T	T	T	T	T	T	T	T	T
11	3707	C	C	C	T	T	T	T	T	C	T
12	3979	A	A	A	A	A	A	A	A	A	A
13	3997	C	C	C	C	C	C	C	C	C	C
14	4140	G	G	G	A	A	A	A	G	A	A
15	4214	C	T	T	C	C	C	T	C	C	C
16	4280	C	C	C	C	C	T	C	C	C	C
17	4313	G	G	G	A	G	A	G	G	G	G
18	4617	T	C	C	C	C	C	C	C	C	C
19	4618	G	G	G	G	G	G	G	G	G	G
20	4992	C	C	T	C	C	C	C	C	C	C

PS		Haplotype Number(c) (Part 3)		
No.(a)	Position(b)	21	22	23
1	2690	C	C	T
2	2864	G	G	G
3	2882	A	G	A
4	2910	T	C	C
5	2949	C	C	C
6	2980	G	G	G
7	2996	T	C	C
8	3259	T	T	T
9	3470	T	C	C
10	3672	C	C	T
11	3707	C	C	C
12	3979	A	A	A
13	3997	C	C	C
14	4140	A	A	A
15	4214	C	C	C
16	4280	C	C	C
17	4313	G	G	G
18	4617	C	C	C
19	4618	G	G	G
20	4992	C	C	C

(a) PS = polymorphic site;

(b) Position of PS within SEQ ID NO:1;

(c) Alleles for haplotypes are presented 5' to 3' in each column;

85

and wherein the haplotype pair is selected from the haplotype pairs shown in the table immediately below, wherein each of the FY haplotype pairs consists of first and second haplotypes which comprise first and second sequences of polymorphisms whose positions in SEQ ID NO:1 and identities are set forth in the table immediately below:

90

PS		Haplotype Pair(c) (Part 1)								
No.(a)	Position(b)	10/10	21/21	7/7	14/14	5/5	12/12	10/16	7/1	
95	1	2690	C/C	C/C	C/C	C/C	C/C	C/C	C/C	
	2	2864	G/G	G/G	G/G	G/G	G/G	G/G	G/A	
	3	2882	A/A	A/A	A/A	A/A	A/A	A/A	A/A	
	4	2910	C/C	T/T	C/C	C/C	C/C	C/C	C/C	
	5	2949	C/C	C/C	C/C	C/C	C/C	C/C	C/C	
100	6	2980	G/G	G/G	G/G	G/G	G/G	G/G	G/G	
	7	2996	C/C	T/T	C/C	C/C	C/C	C/C	C/C	
	8	3259	T/T	T/T	T/T	T/T	T/T	T/T	T/T	
	9	3470	T/T	T/T	T/T	T/T	C/C	T/T	T/T	
	10	3672	T/T	C/C	C/C	T/T	T/T	T/T	C/C	
105	11	3707	C/C	C/C	C/C	T/T	C/C	C/C	C/C	
	12	3979	A/A	A/A	G/G	A/A	A/A	A/A	G/G	
	13	3997	C/C	C/C	C/C	C/C	C/C	C/C	C/C	
	14	4140	G/G	A/A	A/A	A/A	A/A	G/G	A/A	
	15	4214	C/C	C/C	C/C	C/C	C/C	T/T	C/C	
110	16	4280	C/C	C/C	C/C	C/C	C/C	C/C	C/C	
	17	4313	G/G	G/G	G/G	A/A	G/G	G/G	G/A	
	18	4617	C/C	C/C	C/C	C/C	C/C	C/C	C/C	
	19	4618	G/G	G/G	G/G	G/G	G/G	G/G	G/G	
	20	4992	C/C	C/C	C/C	C/C	C/C	C/C	C/C	
115	PS	PS	Haplotype Pair(c) (Part 2)							
	No.(a)	Position(b)	10/14	21/15	5/8	5/14	10/6	18/2	15/19	21/7
120	1	2690	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	2	2864	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	3	2882	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
	4	2910	C/C	T/C	C/C	C/C	C/C	C/C	C/T	T/C
	5	2949	C/C	C/C	C/C	C/C	C/C	C/A	C/C	C/C
125	6	2980	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	7	2996	C/C	T/C	C/C	C/C	C/C	C/C	C/C	T/C
	8	3259	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
	9	3470	T/T	T/T	C/T	C/T	T/T	T/T	T/T	T/T
	10	3672	T/T	C/T	T/C	T/T	T/C	T/T	T/T	C/C
130	11	3707	C/T	C/T	C/C	C/T	C/C	T/T	T/C	C/C
	12	3979	A/A	A/A	A/G	A/A	A/A	A/A	A/A	A/G
	13	3997	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/C
	14	4140	G/A	A/A	A/A	A/A	G/G	G/A	A/A	A/A
	15	4214	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
135	16	4280	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	17	4313	G/A	G/G	G/G	G/A	G/G	G/A	G/G	G/G
	18	4617	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	19	4618	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	20	4992	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C

	PS		Haplotype Pair(c) (Part 3)							
	No.(a)	Position(b)	5/21	10/13	10/1	14/1	10/21	10/18	18/15	10/7
140	1	2690	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	2	2864	G/G	G/G	G/A	G/A	G/G	G/G	G/G	G/G
	3	2882	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
	4	2910	C/T	C/C	C/C	C/C	C/T	C/C	C/C	C/C
145	5	2949	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	6	2980	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	7	2996	C/T	C/C	C/C	C/C	C/T	C/C	C/C	C/C
	8	3259	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
	9	3470	C/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
150	10	3672	T/C	T/T	T/C	T/C	T/C	T/T	T/T	T/C
	11	3707	C/C	C/C	C/C	T/C	C/C	C/T	T/T	C/C
	12	3979	A/A	A/A	A/G	A/G	A/A	A/A	A/A	A/G
	13	3997	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	14	4140	A/A	G/G	G/A	A/A	G/A	G/G	G/A	G/A
155	15	4214	C/C	C/T	C/C	C/C	C/C	C/C	C/C	C/C
	16	4280	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	17	4313	G/G	G/G	G/G	A/G	G/G	G/G	G/G	G/G
	18	4617	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	19	4618	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
160	20	4992	C/C	C/T	C/C	C/C	C/C	C/C	C/C	C/C

	PS		Haplotype Pair(c) (Part 4)							
	No.(a)	Position(b)	5/23	10/5	10/12	5/22	5/7	7/22	21/14	10/3
165	1	2690	C/T	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	2	2864	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	3	2882	A/A	A/A	A/A	A/G	A/A	A/G	A/A	A/A
	4	2910	C/C	C/C	C/C	C/C	C/C	C/C	T/C	C/C
	5	2949	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
170	6	2980	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/C
	7	2996	C/C	C/C	C/C	C/C	C/C	C/C	T/C	C/C
	8	3259	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
	9	3470	C/C	T/C	T/T	C/C	C/T	T/C	T/T	T/T
	10	3672	T/T	T/T	T/T	T/C	T/C	C/C	C/T	T/T
175	11	3707	C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/C
	12	3979	A/A	A/A	A/A	A/A	A/G	G/A	A/A	A/A
	13	3997	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	14	4140	A/A	G/A	G/G	A/A	A/A	A/A	A/A	G/G
	15	4214	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/C
180	16	4280	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	17	4313	G/G	G/G	G/G	G/G	G/G	G/G	G/A	G/G
	18	4617	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	19	4618	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	20	4992	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C

185

PS		Haplotype Pair(c) (Part 5)								
No.(a)	Position(b)	10/9	10/11	20/2	10/20	20/7	10/17	10/4	5/18	
190	1	2690	C/C	C/C	C/C	C/C	C/C	C/C	C/C	
	2	2864	G/G	G/G	G/G	G/G	G/G	G/G	G/G	
	3	2882	A/A	A/A	A/A	A/A	A/A	A/A	A/A	
	4	2910	C/C	C/C	T/C	C/T	T/C	C/C	C/C	
	5	2949	C/C	C/C	C/A	C/C	C/C	C/C	C/C	
	6	2980	G/G	G/G	G/G	G/G	G/G	G/G	G/G	
195	7	2996	C/C	C/C	C/C	C/C	C/C	C/C	C/C	
	8	3259	T/T	T/T	T/T	T/T	T/T	T/T	T/T	
	9	3470	T/T	T/T	T/T	T/T	T/T	T/T	C/T	
	10	3672	T/T	T/T	T/T	T/T	T/C	T/T	T/T	
200	11	3707	C/C	C/C	T/T	C/T	T/C	C/T	C/T	
	12	3979	A/A	A/A	A/A	A/A	A/G	A/A	A/A	
	13	3997	C/C	C/C	C/C	C/C	C/C	C/C	C/C	
	14	4140	G/G	G/G	A/A	G/A	A/A	G/A	A/G	
	15	4214	C/C	C/C	C/C	C/C	C/C	C/T	C/C	
	16	4280	C/C	C/C	C/C	C/C	C/C	C/C	C/C	
205	17	4313	G/G	G/G	G/A	G/G	G/G	G/G	G/G	
	18	4617	C/C	C/T	C/C	C/C	C/C	C/C	C/C	
	19	4618	G/A	G/G	G/G	G/G	G/G	G/G	G/G	
	20	4992	C/C	C/C	C/C	C/C	C/C	C/C	C/C	

(a) PS = polymorphic site;

(b) Position of PS in SEQ ID NO:1;

(c) Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5' to 3' as 1st polymorphism/2nd polymorphism in each column;

wherein a higher frequency of the haplotype or haplotype pair in the trait population than in the reference population indicates the trait is associated with the haplotype or haplotype pair.

12. The method of claim 11, wherein the trait is a clinical response to a drug targeting FY or to a drug for treating a condition or disease associated with FY activity.
13. An isolated oligonucleotide designed for detecting a polymorphism in the Duffy blood group (FY) gene at a polymorphic site (PS) selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS10, PS11, PS12, PS13, PS15, PS18, PS19 and PS20, wherein the selected PS have the position and alternative alleles shown in SEQ ID NO:1.
14. The isolated oligonucleotide of claim 13, which is an allele-specific oligonucleotide that specifically hybridizes to an allele of the FY gene at a region containing the polymorphic site.
15. The allele-specific oligonucleotide of claim 14, which comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:4-19, the complements of SEQ ID NOS:4-19, and SEQ ID NOS:20-51.
16. The isolated oligonucleotide of claim 13, which is a primer-extension oligonucleotide.
17. The primer-extension oligonucleotide of claim 16, which comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:52-83.
18. A kit for haplotyping or genotyping the Duffy blood group (FY) gene of an individual, which comprises a set of oligonucleotides designed to haplotype or genotype each of polymorphic sites

(PS) PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS10, PS11, PS12, PS13, PS15, PS18, PS19 and PS20, wherein the selected PS have the position and alternative alleles shown in SEQ ID NO:1.

19. The kit of claim 18, which further comprises oligonucleotides designed to genotype or haplotype each of PS9, PS14, PS16 and PS17, wherein the selected PS have the position and alternative alleles shown in SEQ ID NO:1.
20. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of:
 - (a) a first nucleotide sequence which comprises a Duffy blood group (FY) isogene, wherein the FY isogene is selected from the group consisting of isogenes 1-23 shown in the table immediately below and wherein each of the isogenes comprises the regions of SEQ ID NO:1 shown in the table immediately below and wherein each of the isogenes 1-23 is further defined by the corresponding sequence of polymorphisms whose positions and identities are set forth in the table immediately below; and

Region	PS	PS	Isogene Number(d) (Part 1)									
Examined(a)	No.(b)	Position(c)	1	2	3	4	5	6	7	8	9	10
2486-3779	1	2690	C	C	C	C	C	C	C	C	C	C
2486-3779	2	2864	A	G	G	G	G	G	G	G	G	G
2486-3779	3	2882	A	A	A	A	A	A	A	A	A	A
2486-3779	4	2910	C	C	C	C	C	C	C	C	C	C
2486-3779	5	2949	C	A	C	C	C	C	C	C	C	C
2486-3779	6	2980	G	G	C	G	G	G	G	G	G	G
2486-3779	7	2996	C	C	C	C	C	C	C	C	C	C
2486-3779	8	3259	T	T	T	C	T	T	T	T	T	T
2486-3779	9	3470	T	T	T	T	C	T	T	T	T	T
2486-3779	10	3672	C	T	T	T	T	C	C	C	T	T
2486-3779	11	3707	C	T	C	C	C	C	C	C	C	C
3872-5302	12	3979	G	A	A	A	A	A	G	G	A	A
3872-5302	13	3997	C	C	C	C	C	C	C	T	C	C
3872-5302	14	4140	A	A	G	G	A	G	A	A	G	G
3872-5302	15	4214	C	C	C	C	C	C	C	C	C	C
3872-5302	16	4280	C	C	C	C	C	C	C	C	C	C
3872-5302	17	4313	G	A	G	G	G	G	G	G	G	G
3872-5302	18	4617	C	C	C	C	C	C	C	C	C	C
3872-5302	19	4618	G	G	G	G	G	G	G	G	A	G
3872-5302	20	4992	C	C	C	C	C	C	C	C	C	C

Region	PS	PS	Isogene Number(d) (Part 2)									
Examined(a)	No.(b)	Position(c)	11	12	13	14	15	16	17	18	19	20
2486-3779	1	2690	C	C	C	C	C	C	C	C	C	C
2486-3779	2	2864	G	G	G	G	G	G	G	G	G	G
2486-3779	3	2882	A	A	A	A	A	A	A	A	A	A
2486-3779	4	2910	C	C	C	C	C	C	C	C	T	T
2486-3779	5	2949	C	C	C	C	C	C	C	C	C	C
2486-3779	6	2980	G	G	G	G	G	G	G	G	G	G
2486-3779	7	2996	C	C	C	C	C	C	C	C	C	C
2486-3779	8	3259	T	T	T	T	T	T	T	T	T	T
2486-3779	9	3470	T	T	T	T	T	T	T	T	T	T
2486-3779	10	3672	T	T	T	T	T	T	T	T	T	T
2486-3779	11	3707	C	C	C	T	T	T	T	T	C	T
3872-5302	12	3979	A	A	A	A	A	A	A	A	A	A
3872-5302	13	3997	C	C	C	C	C	C	C	C	C	C
3872-5302	14	4140	G	G	G	A	A	A	A	G	A	A
3872-5302	15	4214	C	T	T	C	C	C	T	C	C	C
3872-5302	16	4280	C	C	C	C	C	T	C	C	C	C
3872-5302	17	4313	G	G	G	A	G	A	G	G	G	G
3872-5302	18	4617	T	C	C	C	C	C	C	C	C	C
3872-5302	19	4618	G	G	G	G	G	G	G	G	G	G
3872-5302	20	4992	C	C	T	C	C	C	C	C	C	C

Region	PS	PS	Isogene Number(d) (Part 3)		
Examined(a)	No.(b)	Position(c)	21	22	23
2486-3779	1	2690	C	C	T
2486-3779	2	2864	G	G	G
2486-3779	3	2882	A	G	A
2486-3779	4	2910	T	C	C
2486-3779	5	2949	C	C	C
2486-3779	6	2980	G	G	G
2486-3779	7	2996	T	C	C
2486-3779	8	3259	T	T	T
2486-3779	9	3470	T	C	C
2486-3779	10	3672	C	C	T
2486-3779	11	3707	C	C	C
3872-5302	12	3979	A	A	A
3872-5302	13	3997	C	C	C
3872-5302	14	4140	A	A	A
3872-5302	15	4214	C	C	C
3872-5302	16	4280	C	C	C
3872-5302	17	4313	G	G	G
3872-5302	18	4617	C	C	C
3872-5302	19	4618	G	G	G
3872-5302	20	4992	C	C	C

(a) Region examined represents the nucleotide positions defining the start and stop positions within the 1st SEQ ID NO of the sequenced region.

(b) PS = polymorphic site;

(c) Position of PS in SEQ ID NO:1;

(d) Alleles for isogenes are presented 5' to 3' in each column;

(b) a second nucleotide sequence which is complementary to the first nucleotide sequence.

21. The isolated polynucleotide of claim 20, which is a DNA molecule and comprises both the first

and second nucleotide sequences and further comprises expression regulatory elements operably linked to the first nucleotide sequence.

22. A recombinant nonhuman organism transformed or transfected with the isolated polynucleotide of claim 21, wherein the organism expresses a FY protein that is encoded by the first nucleotide sequence.
23. The recombinant nonhuman organism of claim 22, which is a transgenic animal.
24. An isolated fragment of a Duffy blood group (FY) isogene, wherein the fragment comprises at least 10 nucleotides in one of the regions of SEQ ID NO:1 shown in the table immediately below and wherein the fragment comprises one or more polymorphisms selected from the group consisting of thymine at PS1, adenine at PS2, guanine at PS3, thymine at PS4, adenine at PS5, cytosine at PS6, thymine at PS7, cytosine at PS8, thymine at PS10, thymine at PS11, guanine at PS12, thymine at PS13, thymine at PS15, thymine at PS18, adenine at PS19 and thymine at PS20, wherein the selected polymorphism has the position set forth in the table immediately below:

10	Region	PS	PS	Isogene Number(d) (Part 1)									
	Examined(a)	No.(b)	Position(c)	1	2	3	4	5	6	7	8	9	10
	2486-3779	1	2690	C	C	C	C	C	C	C	C	C	C
	2486-3779	2	2864	A	G	G	G	G	G	G	G	G	G
	2486-3779	3	2882	A	A	A	A	A	A	A	A	A	A
15	2486-3779	4	2910	C	C	C	C	C	C	C	C	C	C
	2486-3779	5	2949	C	A	C	C	C	C	C	C	C	C
	2486-3779	6	2980	G	G	C	G	G	G	G	G	G	G
	2486-3779	7	2996	C	C	C	C	C	C	C	C	C	C
	2486-3779	8	3259	T	T	T	C	T	T	T	T	T	T
20	2486-3779	9	3470	T	T	T	T	C	T	T	T	T	T
	2486-3779	10	3672	C	T	T	T	T	C	C	C	T	T
	2486-3779	11	3707	C	T	C	C	C	C	C	C	C	C
	3872-5302	12	3979	G	A	A	A	A	A	G	G	A	A
	3872-5302	13	3997	C	C	C	C	C	C	C	T	C	C
25	3872-5302	14	4140	A	A	G	G	A	G	A	A	G	G
	3872-5302	15	4214	C	C	C	C	C	C	C	C	C	C
	3872-5302	16	4280	C	C	C	C	C	C	C	C	C	C
	3872-5302	17	4313	G	A	G	G	G	G	G	G	G	G
	3872-5302	18	4617	C	C	C	C	C	C	C	C	C	C
30	3872-5302	19	4618	G	G	G	G	G	G	G	G	A	G
	3872-5302	20	4992	C	C	C	C	C	C	C	C	C	C

	Region	PS	PS	Isogene Number(d) (Part 2)									
				11	12	13	14	15	16	17	18	19	20
35	Examined(a)	No.(b)	Position(c)										
	2486-3779	1	2690	C	C	C	C	C	C	C	C	C	C
	2486-3779	2	2864	G	G	G	G	G	G	G	G	G	G
	2486-3779	3	2882	A	A	A	A	A	A	A	A	A	A
40	2486-3779	4	2910	C	C	C	C	C	C	C	C	T	T
	2486-3779	5	2949	C	C	C	C	C	C	C	C	C	C
	2486-3779	6	2980	G	G	G	G	G	G	G	G	G	G
	2486-3779	7	2996	C	C	C	C	C	C	C	C	C	C
45	2486-3779	8	3259	T	T	T	T	T	T	T	T	T	T
	2486-3779	9	3470	T	T	T	T	T	T	T	T	T	T
	2486-3779	10	3672	T	T	T	T	T	T	T	T	T	T
	2486-3779	11	3707	C	C	C	T	T	T	T	T	C	T
50	3872-5302	12	3979	A	A	A	A	A	A	A	A	A	A
	3872-5302	13	3997	C	C	C	C	C	C	C	C	C	C
	3872-5302	14	4140	G	G	G	A	A	A	A	G	A	A
	3872-5302	15	4214	C	T	T	C	C	C	T	C	C	C
55	3872-5302	16	4280	C	C	C	C	C	T	C	C	C	C
	3872-5302	17	4313	G	G	G	A	G	A	G	G	G	G
	3872-5302	18	4617	T	C	C	C	C	C	C	C	C	C
	3872-5302	19	4618	G	G	G	G	G	G	G	G	G	G
	3872-5302	20	4992	C	C	T	C	C	C	C	C	C	C

	Region	PS	PS	Isogene Number(d) (Part 3)		
				21	22	23
60	Examined(a)	No.(b)	Position(c)			
	2486-3779	1	2690	C	C	T
	2486-3779	2	2864	G	G	G
	2486-3779	3	2882	A	G	A
65	2486-3779	4	2910	T	C	C
	2486-3779	5	2949	C	C	C
	2486-3779	6	2980	G	G	G
	2486-3779	7	2996	T	C	C
70	2486-3779	8	3259	T	T	T
	2486-3779	9	3470	T	C	C
	2486-3779	10	3672	C	C	T
	2486-3779	11	3707	C	C	C
75	3872-5302	12	3979	A	A	A
	3872-5302	13	3997	C	C	C
	3872-5302	14	4140	A	A	A
	3872-5302	15	4214	C	C	C
80	3872-5302	16	4280	C	C	C
	3872-5302	17	4313	G	G	G
	3872-5302	18	4617	C	C	C
	3872-5302	19	4618	G	G	G
	3872-5302	20	4992	C	C	C

(a) Region examined represents the nucleotide positions defining the start and stop positions within SEQ ID NO:1 of the regions sequenced;

(b) PS = polymorphic site;

(c) Position of PS within SEQ ID NO:1;

(d) Alleles for FY isogenes are presented 5' to 3' in each column.

25. An isolated polynucleotide comprising a coding sequence for a FY isogene, wherein the coding sequence comprises the regions of SEQ ID NO:2, except at each of the polymorphic sites which

have the positions in SEQ ID NO:2 and polymorphisms set forth in the table immediately below:

PS	PS	Isogene Coding Sequence Number(c) (Part 1)									
No.(a)	Position(b)	2c	3c	4c	6c	9c	10c	11c	12c	13c	14c
14	131	A	G	G	G	G	G	G	G	G	A
15	205	C	C	C	C	C	C	C	T	T	C
16	271	C	C	C	C	C	C	C	C	C	C
17	304	A	G	G	G	G	G	G	G	G	A
18	608	C	C	C	C	C	C	T	C	C	C
19	609	G	G	G	G	A	G	G	G	G	G
20	983	C	C	C	C	C	C	C	C	T	C

PS	PS	Isogene Coding Sequence Number(c) (Part 2)		
No.(a)	Position(b)	16c	17c	18c
14	131	A	A	G
15	205	C	T	C
16	271	T	C	C
17	304	A	G	G
18	608	C	C	C
19	609	G	G	G
20	983	C	C	C

(a) PS = polymorphic site;

(b) Position of PS in SEQ ID NO:2;

(c) Alleles for the isogene coding sequence are presented 5' to 3' in each column; the numerical portion of the isogene coding sequence number represents the number of the parent full FY isogene.

26. A recombinant nonhuman organism transformed or transfected with the isolated polynucleotide of claim 25, wherein the organism expresses a Duffy blood group (FY) protein that is encoded by the polymorphic variant sequence.
27. The recombinant nonhuman organism of claim 26, which is a transgenic animal.
28. An isolated fragment of a FY coding sequence, wherein the fragment comprises one or more polymorphisms selected from the group consisting of thymine at a position corresponding to nucleotide 205, thymine at a position corresponding to nucleotide 608, adenine at a position corresponding to nucleotide 609 and thymine at a position corresponding to nucleotide 983 in SEQ ID NO:2.
29. An isolated polypeptide comprising an amino acid sequence which is a polymorphic variant of a reference sequence for the Duffy blood group (FY) protein, wherein the reference sequence comprises SEQ ID NO:3, except the polymorphic variant comprises one or more variant amino acids selected from the group consisting of phenylalanine at a position corresponding to amino acid position 69, isoleucine at a position corresponding to amino acid position 203, isoleucine at a position corresponding to amino acid position 203 and phenylalanine at a position corresponding to amino acid position 328.
30. An isolated monoclonal antibody specific for and immunoreactive with the isolated polypeptide of claim 29.

31. A method for screening for drugs targeting the isolated polypeptide of claim 29 which comprises contacting the FY polymorphic variant with a candidate agent and assaying for binding activity.
32. An isolated fragment of a FY protein, wherein the fragment comprises one or more variant amino acids selected from the group consisting of phenylalanine at a position corresponding to amino acid position 69, isoleucine at a position corresponding to amino acid position 203, isoleucine at a position corresponding to amino acid position 203 and phenylalanine at a position corresponding to amino acid position 328 in SEQ ID NO:3.
33. A computer system for storing and analyzing polymorphism data for the Duffy blood group gene, comprising:
- a central processing unit (CPU);
 - a communication interface;
 - a display device;
 - an input device; and
 - a database containing the polymorphism data;

wherein the polymorphism data comprises any one or more of the haplotypes set forth in the table immediately below:

10	PS No.(a)	PS Position(b)	Haplotype Number(c) (Part 1)									
			1	2	3	4	5	6	7	8	9	10
	1	2690	C	C	C	C	C	C	C	C	C	C
	2	2864	A	G	G	G	G	G	G	G	G	G
	3	2882	A	A	A	A	A	A	A	A	A	A
15	4	2910	C	C	C	C	C	C	C	C	C	C
	5	2949	C	A	C	C	C	C	C	C	C	C
	6	2980	G	G	C	G	G	G	G	G	G	G
	7	2996	C	C	C	C	C	C	C	C	C	C
	8	3259	T	T	T	C	T	T	T	T	T	T
20	9	3470	T	T	T	T	C	T	T	T	T	T
	10	3672	C	T	T	T	T	C	C	C	T	T
	11	3707	C	T	C	C	C	C	C	C	C	C
	12	3979	G	A	A	A	A	A	G	G	A	A
	13	3997	C	C	C	C	C	C	C	T	C	C
25	14	4140	A	A	G	G	A	G	A	A	G	G
	15	4214	C	C	C	C	C	C	C	C	C	C
	16	4280	C	C	C	C	C	C	C	C	C	C
	17	4313	G	A	G	G	G	G	G	G	G	G
	18	4617	C	C	C	C	C	C	C	C	C	C
30	19	4618	G	G	G	G	G	G	G	G	A	G
	20	4992	C	C	C	C	C	C	C	C	C	C

PS		Haplotype Number(c) (Part 2)									
No.(a)	Position(b)	11	12	13	14	15	16	17	18	19	20
35	1	2690	C	C	C	C	C	C	C	C	C
	2	2864	G	G	G	G	G	G	G	G	G
	3	2882	A	A	A	A	A	A	A	A	A
	4	2910	C	C	C	C	C	C	C	T	T
40	5	2949	C	C	C	C	C	C	C	C	C
	6	2980	G	G	G	G	G	G	G	G	G
	7	2996	C	C	C	C	C	C	C	C	C
	8	3259	T	T	T	T	T	T	T	T	T
	9	3470	T	T	T	T	T	T	T	T	T
45	10	3672	T	T	T	T	T	T	T	T	T
	11	3707	C	C	C	C	C	C	C	C	C
	12	3979	A	A	A	A	A	A	A	A	A
	13	3997	C	C	C	C	C	C	C	C	C
	14	4140	G	G	G	A	A	A	G	A	A
50	15	4214	C	T	T	C	C	C	T	C	C
	16	4280	C	C	C	C	C	T	C	C	C
	17	4313	G	G	G	A	G	A	G	G	G
	18	4617	T	C	C	C	C	C	C	C	C
	19	4618	G	G	G	G	G	G	G	G	G
55	20	4992	C	C	T	C	C	C	C	C	C

PS		Haplotype Number(c) (Part 3)			
No.(a)	Position(b)	21	22	23	
60	1	2690	C	C	T
	2	2864	G	G	G
	3	2882	A	G	A
	4	2910	T	C	C
	5	2949	C	C	C
65	6	2980	G	G	G
	7	2996	T	C	C
	8	3259	T	T	T
	9	3470	T	C	C
	10	3672	C	C	T
70	11	3707	C	C	C
	12	3979	A	A	A
	13	3997	C	C	C
	14	4140	A	A	A
	15	4214	C	C	C
75	16	4280	C	C	C
	17	4313	G	G	G
	18	4617	C	C	C
	19	4618	G	G	G
	20	4992	C	C	C

80

(a) PS = polymorphic site;

(b) Position of PS within SEQ ID NO:1;

(c) Alleles for haplotypes are presented 5' to 3' in each column;

the haplotype pairs set forth in the table immediately below:

85

	PS		Haplotype Pair(c) (Part 1)							
	No.(a)	Position(b)	10/10	21/21	7/7	14/14	5/5	12/12	10/16	7/1
90	1	2690	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	2	2864	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A
	3	2882	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
	4	2910	C/C	T/T	C/C	C/C	C/C	C/C	C/C	C/C
	5	2949	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	6	2980	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
95	7	2996	C/C	T/T	C/C	C/C	C/C	C/C	C/C	C/C
	8	3259	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
	9	3470	T/T	T/T	T/T	T/T	C/C	T/T	T/T	T/T
	10	3672	T/T	C/C	C/C	T/T	T/T	T/T	T/T	C/C
100	11	3707	C/C	C/C	C/C	T/T	C/C	C/C	C/T	C/C
	12	3979	A/A	A/A	G/G	A/A	A/A	A/A	A/A	G/G
	13	3997	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	14	4140	G/G	A/A	A/A	A/A	A/A	G/G	G/A	A/A
	15	4214	C/C	C/C	C/C	C/C	C/C	T/T	C/C	C/C
	16	4280	C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/C
105	17	4313	G/G	G/G	G/G	A/A	G/G	G/G	G/A	G/G
	18	4617	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	19	4618	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	20	4992	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
110	PS		Haplotype Pair(c) (Part 2)							
	No.(a)	Position(b)	10/14	21/15	5/8	5/14	10/6	18/2	15/19	21/7
115	1	2690	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	2	2864	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	3	2882	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
	4	2910	C/C	T/C	C/C	C/C	C/C	C/C	C/T	T/C
	5	2949	C/C	C/C	C/C	C/C	C/C	C/A	C/C	C/C
	6	2980	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
120	7	2996	C/C	T/C	C/C	C/C	C/C	C/C	C/C	T/C
	8	3259	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
	9	3470	T/T	T/T	C/T	C/T	T/T	T/T	T/T	T/T
	10	3672	T/T	C/T	T/C	T/T	T/C	T/T	T/T	C/C
	11	3707	C/T	C/T	C/C	C/T	C/C	T/T	T/C	C/C
	12	3979	A/A	A/A	A/G	A/A	A/A	A/A	A/A	A/G
125	13	3997	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/C
	14	4140	G/A	A/A	A/A	A/A	G/G	G/A	A/A	A/A
	15	4214	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	16	4280	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
130	17	4313	G/A	G/G	G/G	G/A	G/G	G/A	G/G	G/G
	18	4617	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	19	4618	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	20	4992	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C

	PS		Haplotype Pair(c) (Part 3)							
	No.(a)	Position(b)	5/21	10/13	10/1	14/1	10/21	10/18	18/15	10/7
135	1	2690	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	2	2864	G/G	G/G	G/A	G/A	G/G	G/G	G/G	G/G
	3	2882	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
	4	2910	C/T	C/C	C/C	C/C	C/T	C/C	C/C	C/C
140	5	2949	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	6	2980	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	7	2996	C/T	C/C	C/C	C/C	C/T	C/C	C/C	C/C
	8	3259	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
	9	3470	C/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
145	10	3672	T/C	T/T	T/C	T/C	T/C	T/T	T/T	T/C
	11	3707	C/C	C/C	C/C	T/C	C/C	C/T	T/T	C/C
	12	3979	A/A	A/A	A/G	A/G	A/A	A/A	A/A	A/G
	13	3997	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	14	4140	A/A	G/G	G/A	A/A	G/A	G/G	G/A	G/A
150	15	4214	C/C	C/T	C/C	C/C	C/C	C/C	C/C	C/C
	16	4280	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	17	4313	G/G	G/G	G/G	A/G	G/G	G/G	G/G	G/G
	18	4617	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	19	4618	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
155	20	4992	C/C	C/T	C/C	C/C	C/C	C/C	C/C	C/C

	PS		Haplotype Pair(c) (Part 4)							
	No.(a)	Position(b)	5/23	10/5	10/12	5/22	5/7	7/22	21/14	10/3
	1	2690	C/T	C/C	C/C	C/C	C/C	C/C	C/C	C/C
160	2	2864	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	3	2882	A/A	A/A	A/A	A/G	A/A	A/G	A/A	A/A
	4	2910	C/C	C/C	C/C	C/C	C/C	C/C	T/C	C/C
	5	2949	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	6	2980	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/C
165	7	2996	C/C	C/C	C/C	C/C	C/C	C/C	T/C	C/C
	8	3259	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
	9	3470	C/C	T/C	T/T	C/C	C/T	T/C	T/T	T/T
	10	3672	T/T	T/T	T/T	T/C	T/C	C/C	C/T	T/T
	11	3707	C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/C
170	12	3979	A/A	A/A	A/A	A/A	A/G	G/A	A/A	A/A
	13	3997	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	14	4140	A/A	G/A	G/G	A/A	A/A	A/A	A/A	G/G
	15	4214	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/C
	16	4280	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
175	17	4313	G/G	G/G	G/G	G/G	G/G	G/G	G/A	G/G
	18	4617	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	19	4618	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	20	4992	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C

180

PS		Haplotype Pair(c) (Part 5)								
No.(a)	Position(b)	10/9	10/11	20/2	10/20	20/7	10/17	10/4	5/18	
185	1	2690	C/C	C/C	C/C	C/C	C/C	C/C	C/C	
	2	2864	G/G	G/G	G/G	G/G	G/G	G/G	G/G	
	3	2882	A/A	A/A	A/A	A/A	A/A	A/A	A/A	
	4	2910	C/C	C/C	T/C	C/T	T/C	C/C	C/C	
	5	2949	C/C	C/C	C/A	C/C	C/C	C/C	C/C	
	6	2980	G/G	G/G	G/G	G/G	G/G	G/G	G/G	
	7	2996	C/C	C/C	C/C	C/C	C/C	C/C	C/C	
190	8	3259	T/T	T/T	T/T	T/T	T/T	T/C	T/T	
	9	3470	T/T	T/T	T/T	T/T	T/T	T/T	C/T	
	10	3672	T/T	T/T	T/T	T/T	T/C	T/T	T/T	
195	11	3707	C/C	C/C	T/T	C/T	T/C	C/T	C/T	
	12	3979	A/A	A/A	A/A	A/A	A/G	A/A	A/A	
	13	3997	C/C	C/C	C/C	C/C	C/C	C/C	C/C	
	14	4140	G/G	G/G	A/A	G/A	A/A	G/A	A/G	
	15	4214	C/C	C/C	C/C	C/C	C/C	C/T	C/C	
	16	4280	C/C	C/C	C/C	C/C	C/C	C/C	C/C	
	17	4313	G/G	G/G	G/A	G/G	G/G	G/G	G/G	
200	18	4617	C/C	C/T	C/C	C/C	C/C	C/C	C/C	
	19	4618	G/A	G/G	G/G	G/G	G/G	G/G	G/G	
	20	4992	C/C	C/C	C/C	C/C	C/C	C/C	C/C	

(a) PS = polymorphic site;

(b) Position of PS in SEQ ID NO:1;

(c) Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5' to 3' as 1st polymorphism/2nd polymorphism in each column;

and the frequency data in Tables 5 and 6.

34. A genome anthology for the Duffy blood group (FY) gene which comprises two or more FY isogenes selected from the group consisting of isogenes 1-23 shown in the table immediately below, and wherein each of the isogenes comprises the regions of SEQ ID NO:1 shown in the table immediately below and wherein each of the isogenes 1-23 is further defined by the corresponding sequence of polymorphisms whose positions and identities are set forth in the table immediately below:

	Region Examined(a)	PS No.(b)	PS Position(c)	Isogene Number(d) (Part 1)									
				1	2	3	4	5	6	7	8	9	10
10	2486-3779	1	2690	C	C	C	C	C	C	C	C	C	C
	2486-3779	2	2864	A	G	G	G	G	G	G	G	G	G
	2486-3779	3	2882	A	A	A	A	A	A	A	A	A	A
	2486-3779	4	2910	C	C	C	C	C	C	C	C	C	C
	2486-3779	5	2949	C	A	C	C	C	C	C	C	C	C
15	2486-3779	6	2980	G	G	C	G	G	G	G	G	G	G
	2486-3779	7	2996	C	C	C	C	C	C	C	C	C	C
	2486-3779	8	3259	T	T	T	C	T	T	T	T	T	T
	2486-3779	9	3470	T	T	T	T	C	T	T	T	T	T
	2486-3779	10	3672	C	T	T	T	T	C	C	C	T	T
20	2486-3779	11	3707	C	T	C	C	C	C	C	C	C	C
	3872-5302	12	3979	G	A	A	A	A	A	G	G	A	A
	3872-5302	13	3997	C	C	C	C	C	C	C	T	C	C
	3872-5302	14	4140	A	A	G	G	A	G	A	A	G	G
	3872-5302	15	4214	C	C	C	C	C	C	C	C	C	C
25	3872-5302	16	4280	C	C	C	C	C	C	C	C	C	C
	3872-5302	17	4313	G	A	G	G	G	G	G	G	G	G
	3872-5302	18	4617	C	C	C	C	C	C	C	C	C	C
	3872-5302	19	4618	G	G	G	G	G	G	G	G	A	G
	3872-5302	20	4992	C	C	C	C	C	C	C	C	C	C
30	Region Examined(a)	PS No.(b)	PS Position(c)	Isogene Number(d) (Part 2)									
				11	12	13	14	15	16	17	18	19	20
35	2486-3779	1	2690	C	C	C	C	C	C	C	C	C	C
	2486-3779	2	2864	G	G	G	G	G	G	G	G	G	G
	2486-3779	3	2882	A	A	A	A	A	A	A	A	A	A
	2486-3779	4	2910	C	C	C	C	C	C	C	C	T	T
	2486-3779	5	2949	C	C	C	C	C	C	C	C	C	C
40	2486-3779	6	2980	G	G	G	G	G	G	G	G	G	G
	2486-3779	7	2996	C	C	C	C	C	C	C	C	C	C
	2486-3779	8	3259	T	T	T	T	T	T	T	T	T	T
	2486-3779	9	3470	T	T	T	T	T	T	T	T	T	T
	2486-3779	10	3672	T	T	T	T	T	T	T	T	T	T
45	2486-3779	11	3707	C	C	C	T	T	T	T	T	C	T
	3872-5302	12	3979	A	A	A	A	A	A	A	A	A	A
	3872-5302	13	3997	C	C	C	C	C	C	C	C	C	C
	3872-5302	14	4140	G	G	G	A	A	A	A	G	A	A
	3872-5302	15	4214	C	T	T	C	C	C	T	C	C	C
50	3872-5302	16	4280	C	C	C	C	C	T	C	C	C	C
	3872-5302	17	4313	G	G	G	A	G	A	G	G	G	G
	3872-5302	18	4617	T	C	C	C	C	C	C	C	C	C
	3872-5302	19	4618	G	G	G	G	G	G	G	G	G	G
	3872-5302	20	4992	C	C	T	C	C	C	C	C	C	C

55

55	Region Examined(a)	PS No.(b)	PS Position(c)	Isogene Number(d) (Part 3)		
				21	22	23
	2486-3779	1	2690	C	C	T
	2486-3779	2	2864	G	G	G
	2486-3779	3	2882	A	G	A
60	2486-3779	4	2910	T	C	C
	2486-3779	5	2949	C	C	C
	2486-3779	6	2980	G	G	G
	2486-3779	7	2996	T	C	C
	2486-3779	8	3259	T	T	T
65	2486-3779	9	3470	T	C	C
	2486-3779	10	3672	C	C	T
	2486-3779	11	3707	C	C	C
	3872-5302	12	3979	A	A	A
	3872-5302	13	3997	C	C	C
70	3872-5302	14	4140	A	A	A
	3872-5302	15	4214	C	C	C
	3872-5302	16	4280	C	C	C
	3872-5302	17	4313	G	G	G
	3872-5302	18	4617	C	C	C
75	3872-5302	19	4618	G	G	G
	3872-5302	20	4992	C	C	C

(a) Region examined represents the nucleotide positions defining the start and stop positions within SEQ ID NO:1 of the regions sequenced;

(b) PS = polymorphic site;

(c) Position of PS within SEQ ID NO:1;

(d) Alleles for FY isogenes are presented 5' to 3' in each column.

POLYMORPHISMS IN THE FY GENE

GGCAAAGGTT	GGGAGTGGCT	TTTCCTCTGG	TAGCCACACA	CCTGAGCACT	
ACGGACAGGG	AGGCAGGTGC	CACCTTGACA	CCTCTCTTCC	ATAGCAATGG	100
GAAAGTGATG	AGTGCGGGAG	TCCTGAGGAG	ATGTGGCCTG	CAGACAACAT	
GCAGCCATGC	AGGGACCCAG	GACTGTAACC	TGGGGAGGAC	GCGGGTCCCT	200
GCAAGGAAGA	GTAAGTTTGG	AGAGGAAGGA	TGGAGGTGGA	CTCTCACCCC	
ATTCCCCCG	GAAATGAACA	AAGCCGGGCC	CTTTCCATAG	GAAGTGCCTT	300
TGGAGATAGC	AGAGTGTGGC	TGCCCCCTCT	TGCTCCAGCA	GCAGTGGGAG	
AGGCACTGCT	CTGGGGCCTG	AACTGCCTCT	GCTTCCCCCC	CTGAGGGGCC	400
CCTCACTCTT	ACCCAAGACT	CTGGATTGTT	GCACGGCAAC	CACTCCTCCC	
ATGGCATTGC	TCAGCAACTA	CTTCTCCCTT	CCCGGCCACC	CTGTGCCCCC	500
TTCTCTGGTCC	CAACGCCAGC	CCTTCATCCT	TCCTCCCTCA	GCAGCCAGGC	
AGACATAACA	ACAAAACCTAC	TAAAAGGAGC	TTCACTGCAG	TGAGCTGTTT	600
CCTGCCCAAA	CTAAGGGAAT	AATGTGAAC	GTGTGCATGT	GTGTGGTGTG	
TATGCATGTG	TGCATGTGTG	TGTGTGTGTG	TGCATGTGTG	TGAGTGAGTG	700
AGAGGCAGAG	CGAGGAACTG	AGGAGGAGGG	CTAAGAGCCA	GGGGTCTCTG	
GCAAGTGGAC	AGGGCTGTGG	GACATGTTGG	GGAGGCTTTG	GGAATGGGGT	800
ATTCTAGTC	AGGGTTTACA	CCTCACCTGG	GATGTTGTTT	CATGCTGGTA	
TTTCCTCTGC	CACCCCCAAT	GCCCCATCGT	CTTGGAGAAA	GGAGTCCCCG	900
GGTGTGTGTT	TGCCCAGCTG	TCCATTCTAT	CTCTCCCTTA	AACACAGAGC	
ATTCAGCCCT	TCCCTGGATT	TCCCTCCTCT	GAGCCATGGA	GTCAGTGCCA	1000
CAGCCTTTGC	TATGCACCTC	TCAGGCCTCT	CCTTGGCGTT	GACCCTGGAA	
AGACCTACCA	CCACCTATTT	TTTCCCATAG	TCTGTACCCA	GTGAGTTGAA	1100
GGCTGGGTCC	CCACCCTTCC	TTTTGATTTC	CTGTCTTCCT	TCTCGTGGCC	
CCAGCTGGTT	GCTGTGGAGA	TGAGGTTTCT	GGTCCCTCCCT	GTCCTGGCTG	1200
GACTGCCCCG	CCTCAGATCC	AGGATGCCCT	TGGCATCGCT	CCCACCCTCC	
CCCAGCTTTT	CCTCCCTGGT	CTGACAATGG	GCATGCAAAA	AGGGGCAGCT	1300
GCAATCTAGC	AGGCCTGCCC	ACCCCTTCA	GTTCAAGGTAA	TACAGTTGTG	
AATCTTCCAG	CCGCTGGTTA	GGGCCTTGGG	CACCACAGGC	AGCCCCTCAC	1400
CTAAGCCGGG	GCCTACTCCT	CTTACAACAG	CAAGAGAGCC	CTGGGGCCCC	
AGGCCTGTTG	AGCTTCTTGT	CTCCCAGCAC	CCGCTTTTGG	GAAAATGACT	1500
TTTCCTCTTC	AAGCTGAACC	ACTCTGTCCA	TATTACACAG	AAGCATATTT	
TGTACGGGGG	GGTGGGAGGG	AGAGGGGCTG	TTGTGCTGTG	TGTGTCTGTC	1600
CAGGGGTGGG	GGGGTGGGGG	AAGGGAGCAG	GGAGGGGACC	GTGTATCTTT	
ATAATCTTTC	TAATCTCCT	GTGCTAATCT	CAGAGGGGTC	ACCCTCAATA	1700
TATCTGGATT	ATCCGTGTCA	TTCAGCTGCC	TCCTTTCTGG	TCCTCTTGCT	
GCTGCTGGGA	TGTGTGTATG	TGAGGGTCTT	CTTCCCATAC	CCCTTGACAC	1800
TGGTGCTGG	TGCCTCAAAA	GGTGGTGTGT	CCCTTGCCAG	GCCACTCTCA	
AGAATATCTA	TGTACAGCAA	CAATATAACT	CTACAAGGGA	GAGAAGTGTG	1900
TTCACTTCCT	TTTGCTAAGC	CCTTCCTTTC	CAGAGAGTGT	CTTGGGGGGC	
ATCTGACTGC	TTCCCCCAC	CCTCTGCCAG	GCATTGCTGG	AGAATGTTAA	2000
GACGGCGATG	GAGATGCCAT	CAACCCACC	CTGCAGAGCA	TCACCAGACA	
CCACCAGACC	AAATTCACCT	TCCAGCCCCT	TCATGTTGAA	CCTGAAACTT	2100
GAGCTAGTGT	CTTGGGAGAA	AAGGGGGAAA	TCTCTACGAG	GTACCCATCC	
TTCTGCACCT	TAGGTCTGAG	GTGCTTGGCC	CCCTAGGAAG	CCCTACATGA	2200
ATGGGACAGA	AGGTCCTTAA	CAACACTGGA	GATGAAGCAG	CCGATGCTGT	
TTTGGACAAA	TGAAACAGCG	TCCCCTAACC	AGCCCTTTCT	ATCTCATTGT	2300
TCTGACTTGG	ACACGCCATG	GCTCACCGCT	CCCAAAGTCC	CCACTATGTC	
TCCCTAGCTG	AGGAAATAAA	AGCAGAGAGG	GGTGATGAAA	CAGTGACGAT	2400
CCTGGGGAAA	CAGCTGAGGA	GGGGAGGGAG	GGGGAAGAAG	CCACTAAAAA	

FIGURE 1A

AGTGAAATGT	GCTTGGGAGA	ATCGGCCTGC	CTGCAGGGTA	GATGCCCTTT	2500
CTCTCTGCTG	GCCAGCTCTG	CCCCTCAGTG	AGAAACTTTA	CATATTGCTA	
AGATGCCTGG	CCAATGAAAC	AGTTCCAGAG	ACTTTATGTC	CCCCAGTAGA	2600
AATATGAATA	GAAATCACCC	TGTGGGCAAT	GGTCCCATT	TAAAATATGC	
TGTCCCATTG	TCCCCTAGAG	CCTACTTTAA	CTTGTCAGAC	CATGTATTCC	2700
			T		
ACTTCATATG	CAAGAGGCAT	GCACTGAGCC	CATAGGTGGC	TAGGCAAACA	
CCCAATAGCT	CCCTGAAATG	GCTTCATTAT	GCAGCCTCGA	CAGCCACCCC	2800
AACCCCTCCA	CTCTCACACT	GAAACACCCA	GACCTAGAGA	TAGCTAGACA	
CACCCAGACA	CCCGCCAAGC	CCCTCACATA	CAGATATGTG	CACAATGATA	2900
	A		G		
CACAGCAAAC	GTACACAGAG	TTCAGTACAC	ACAAAGAGCT	CACGCCCACG	
	T			A	
TGCACACACC	CCTCAGTTGG	GACAGAGTTG	ACCACCACCA	CCTTTCTCCC	3000
		C		T	
AAACACATGG	CTTTTGAAC	GCCTTTCTTT	GGATCCAGTT	CAAGGGGATG	
GAGGAGCAGT	GAGAGTCAGC	CGCCCTTCCA	CTCCAATTT	CCAGCACCTC	3100
CCTTATCTCT	GCCTCACAA	TCACCCAGCC	CCCCCTCTTT	CCTTCCTTGT	
GCTTGAAGAA	TCTCTCCTTG	CTGGAAAGCC	CCCTGTTTT	TCAATCTCCC	3200
TTTCCACTTC	GGTAAAATCT	CTACTTGCTG	GAAAGCCCC	TGTTTTCTCA	
ATCTCCCTTT	CCACTTCGGT	AAAATGCCCA	CTTTCTGGTC	CCCACCTTTT	3300
	C				
TCCTGAGTGT	AGTCCCAACC	AGCCAAATCC	AACCTCAAAA	CAGGAAGACC	
CAAGGCCAGT	GACCCCCATA	GGCCTGAGGC	TTGTGCAGGC	AGTGGGCGTG	3400
GGGTAAGGCT	TCCTGATGCC	CCCTGTCCCT	GCCAGAACC	TGATGGCCCT	
CATTAGTCCT	TGGCTCTTAT	CTTGGAAGCA	CAGGCGTGTA	CAGCCGTCCC	3500
		C			
AGCCCTTCTG	TCTGCGGGCC	TGAACCAAAC	GGTGCCATGG	GGAAGTGTCT	
GCACAGGGTG	AGTATGGGGC	CAGGCCCCAG	AGTCCCTTAT	CCCTATGCCC	3600
CTCATTTCCC	CTGCTGTTTG	CCCCTCAGTC	TTTATATCTC	TTCTTTTCCC	
TCCTCATCTT	TTCTCCCTTC	CCGCTTTTTT	CCTCTTCCTT	CAAAGTCTTT	3700
		T			
TTCTTTCTCT	CCTTCCTATG	CTAGCCTCCT	AGCTCCCTCT	TGTGTCCCTC	
	T				
CCTTTGCCTT	TGAGTCAGTT	CCATCCTGGT	CTCTTGCTGC	CTTTCCTTCT	3800
GACCTTGAC	TGCTCCTCCA	GCCCCAGCTG	CCCTGGCTTC	CCCAGGACTG	
TTCTGCTCC	GGCTCTTCAG	GCTCCCTGCT	TTGTCTTTTT	CCACTGTCCG	3900
CACTGCATCT	GACTCCTGCA	GAGACCTTGT	TCTCCACCC	GACCTTCCTC	
TCTGTCTCC	CCTCCACCT	GCCCCTCAAT	TCCAGGAGA	CTCTTCCGGT	4000
		G		T	
GTAACCTCTGA	TGGCCTCCTC	TGGGTATGTC	CTCCAGGCGG	AGCTCTCCCC	
[exon 1: 4010..					
CTCAACTGAG	AACTCAAGTC	AGCTGGACTT	CGAAGATGTA	TGGAATTCTT	4100
CCTATGGTGT	GAATGATTCC	TTCCCAGATG	GAGACTATGA	TGCCAACCTG	
			G		
GAAGCAGCTG	CCCCCTGCCA	CTCCTGTAAC	CTGCTGGATG	ACTCTGCACT	4200
GCCCTTCTTC	ATCCTCACCA	GTGTCCTGGG	TATCCTAGCT	AGCAGCACTG	
	T				
TCCTCTTCAT	GCTTTTCAGA	CCTCTCTTCC	GCTGGCAGCT	CTGCCCTGGC	4300
		T			
TGGCCTGTCC	TGGCACAGCT	GGCTGTGGGC	AGTGCCCTCT	TCAGCATTGT	
	A				
GGTGCCCGTC	TTGGCCCCAG	GGCTAGGTAG	CACTCGCAGC	TCTGCCCTGT	4400

FIGURE 1B

GTAGCCTGGG	CTACTGTGTC	TGGTATGGCT	CAGCCTTTGC	CCAGGCTTTG	
CTGCTAGGGT	GCCATGCCTC	CCTGGGCCAC	AGACTGGGTG	CAGGCCAGGT	4500
CCCAGGCCCTC	ACCCTGGGGC	TCACTGTGGG	AATTTGGGGA	GTGGCTGCCC	
TACTGACACT	GCCTGTCACC	CTGGCCAGTG	GTGCTTCTGG	TGGACTCTGC	4600
ACCCTGATAT	ACAGCACGGA	GCTGAAGGCT	TTGCAGGCCA	CACACACTGT	
TA					
AGCCTGTCTT	GCCATCTTTG	TCTTGTTGCC	ATTGGGTTTG	TTTGGAGCCA	4700
AGGGGCTGAA	GAAGGCATTG	GGTATGGGGC	CAGGCCCTTG	GATGAATATC	
CTGTGGGCCT	GGTTTATTTT	CTGGTGGCCT	CATGGGGTGG	TTCTAGGACT	4800
GGATTTCCTG	GTGAGGTCCA	AGCTGTTGCT	GTTGTCAACA	TGTCTGGCCC	
AGCAGGCTCT	GGACCTGCTG	CTGAACCTGG	CAGAAGCCCT	GGCAATTTTG	4900
CACTGTGTGG	CTACGCCCTT	GCTCCTCGCC	CTATTCTGCC	ACCAGGCCAC	
CCGCACCCTC	TTGCCCTCTC	TGCCCCCTCC	TGAAGGATGG	TCTTCTCATC	5000
T					
TGGACACCCT	TGGAAGCAAA	TCCTAGTTCT	CTTCCCACCT	GTCAACCTGA	
..50261					
ATTAAAGTCT	ACACTGCCTT	TGTGAAGCGG	GTGGTTTCTT	ATTTTGTCTG	5100
GGGAGAAGAA	GGAGAATGGA	GAGAGAGACA	TTTTTATGTC	AGACTTTCTT	
GCCAGTGTCT	GCTTCTATAG	CTGGCTTGGG	AAGAAGGTGA	ATGATGAATA	5200
AATACCCTCA	GGGTACACAG	ATGTTCTCTT	GAGGTGTGGG	GTCACGGCCA	
TCTCAAGGGA	GAAGAGAAGA	GGAACCAGAG	CATGAGGGGA	GTCATTAAAC	5300
CAAAAAAAG	AGAAGGGATG	GCTTAGCTGG	AAAAAAGCT	GTTCTGGGAA	
GCAAAATGGAA	TAGGAACCTCA	AACTGAGAGA	TAAACAGTGA	AGAGTGATGA	5400
CAAAGCCCAG	AGCAATACCA	CCTCCCCCTG	TCCAACCTGC	CCAGCCTCTG	
TCTTCTGTCT	CCTCTCTGGC	TTTGTTTAGT	GATTAGGACA	GTGGTGGGGA	5500
AGGTGAAAGA	AGCATCCCAG	GGGATGTTAC	TCAGTTCAGG	GAACATATCA	
AGGTAATTTA	AAAAGCCACT	TCCTGGGAGT	CATCTCTCCC	AGGTTCTCTA	5600
GCATGACCTG	AATGTGCGTG	CGTGTGTGTG	TGTGTGTGTG	TGTACACATC	
TGTTTCTCGA	TCTGTTAGAA	TCTACCTTTA	TGTTAGATGT	ATGCATGTAA	5700
AAACATATGT	CCACCCATGA	GCTTGCATCT	CTGTCAGCAC	CTGAACTGCG	
CACACCTGTG	CGTGTGCACT	GACTTTTCTC	AGGACCCAAA	CCCCCACTCA	5800
ATTCTGCACT	CATCCCTGTT	CACAGGATAT	AGAATCGGGA	TTTATGACTC	
ACTCCTTACC	CAAATGAGTT	TTCTTTACCC	TGGTTTTTAA	GCCTAGTCTT	5900
TTCTGTGTAG	GATGTGTGGA	GGGAAGAAAA	GATCAAGAAG	TTGTGAAGGG	
TGGAGAAACT	TGAAGGGGGA	GGCCCTGATT	TGATTCATCT	TCTGCTTGGA	6000
ATTCCCCGAA	TTTCCCTTTC	AGAATCTCAG	CTTTTGAAAT	AAACCTTTAT	
TTCCACACATA	CATCTTTCCCT	TCCACCTTCC	ACACAATACC	CCAATCCCTT	6100
GGGCACCTTT	TTCCCAACCC	CTGATTCTCT	GGCTGCTTAA	TCATGACCTT	
TGAGATTTTT	CTCAGTCTCT	ACCTACCCAA	GTTTAGATGG	CTGGAAGGAC	6200
AGAAACCCCT	CCTCATCAGG	GGCACAGCTT	TTACCACCAA	GAGCAAATTC	
ACCCTCTACC	CAAGAGGCTA	CAAAACAGTT	AGTTCCTACC	TCTAACCCAA	6300
CTAAAGGCTG	GGGAAACTTG	AGCAGATACG	TTCTATCAGT	TTGAACCCAA	
TTACCATCTT	ACCATTTTCC	AAAGATATGC	TATACCTGGT	TTCTTTACTA	6400
AAATGTTTCT	GCTTGACTCT	CTGGGCTTGG	GAATAGTAGG	CGAGTGCGGG	
AGAGGTGCAG	AGATGAGTTA	GAATAGCTTA	GGCAGGAGGG	TGCAAAAGGC	6500
TTAGGGAATT	TTCTTGGGTG	GGTGCCACGA	CAAGGCCCTCT	AAATCTCCCA	
CCTCCTGTCT	CTTAGCAACC	ACCAGGTTAG	CTCCTGATTG	GTTTCGTCCTC	6600
AATTGAAAGG	CGGGATTTAG	GGACCGATTG	AGACGCGGGA	GACATTCTGA	
AACAGAAAGG	AAGGGAGAGA	AAATGAAGAG	AAAGGAAATA	ATTTACAAAC	6700
CTAAATTATG	CTCTGGTTTC	CAACCACAGT	TCATGAATGT	GTTCTAGTAT	
TTTTTCCCCC	GCTTTTTTTT	TTCCAGGCTT	CTCTCAATAT	CCCCCTCCCG	6800
TCCTTGACCA	CTCTTGCAAT	TCTACCAGAT	GTTGCTGTCC	TCCCTTACAA	
GGTACTGATT	TGGAAGCTGA	CCTAGTTGAG	GGGGAGGAGA	GGGCGTTTTT	6900

FIGURE 1C

GACTCCCTGA	ATCTTCCAGT	GTCAACCTGA	TGCAAGGGAG	GCTTAATTTA	
AGACCAGTAG	GCTTGTCTTA	TCTGCCCCCA	ACCCTGTGCC	TCTGGATAGA	7000
AATCCCTGGT	CAGTCAGTCC	AGTTAGAGAG	AACCCAGAC	TCCTGGGTAA	
TAGCTTGGCA	GCTCTCATGG	CTTTCACAAG	GGAAAGGCAG	CTGCAGAAGC	7100
CCGAAGCTGC	TAAGAGGTTA	GGGTGGGCTG	GAGACAGTGC	CCTACCCCCG	
CCCCCTGCTA	CATCCTCCTC	ATCCCCACCC	CCACCGGGAT	TGCTCCAGGC	7200
CTTTTGGGCT	GCCCTTTCCC	TGCCATTACC	TAGGCAGCAC	TTGGAGAGCT	
CCTCCTTAAG	TCTAACCCGG	ACCTCAGTCA	TTTCTTTAAA	GCTTCTTTGG	7300
GGACCTGCCA	CCCCATGCAT	TTAACCCACT	GCATGCCATC	AACCACTCTA	
AAATTGGTCT	GAGTCTGGCA	TCTTTTCTGC	AACCCCTCAG	GAATACAAAT	7400
CCTGTCTCCT	TAAAGCCCTT	AAGAATTTAA	TCTTAGGGTT	GGCAGGGACT	
TTAGCTGTGT	ATGAGATATT	GGGCATCCTA	GCTAAAGAAA	AAAATCCTCT	7500
CAGAAAGATG	AGAGCCAGGG	AAGCAAGCTC	TTGGGAAAAC	ACAGGACCCT	
GAGGAAGGTC	AGTTTGCTTT	GCTTTCTAAA	GGAGAGAGAT	CTATTATTCA	7600
AGGGAAGTTT	GAACATCACA	TTGACGCTCA	TAGTTCATTT	ATTCCAAGCT	
GAGGCCCTC	CCTTAGGATT	TAGAAAACAA	ATACTTGGTC	CTCACACCCT	7700
TTTTCCATTC	CTATTTCCCT	ATCCCCCAAC	CCCATCACCA	CCTTCCTCCC	
TCAGAGGAAT	TCTGATTGAG	AACTTCACTG	GGATTTCAAA	CCCAATTCAT	7800
CGCCAACCTCT	AATTGCCAGA	GATTTGCATG	AAAACCATCG	TATGCTATCT	
AATTATTCTG	ACAACAGCAG	CCCGCCGTCT	GGGCACAAGG	AGAATCGGAG	7900
TTTTAATTAA	CAATAATGCA	CCTTGCTGAC	GAATGCGACT	GTTTAGGTTA	
ATTAACAAGT	CCAAGTCCTT	CCAAATCATC	TCTAGACATC	TAGGTGATTT	8000
GGGCAGGAAG	GGTGTGGGGA	ACACAGGGAG	GGATGGGGAG	TGTTTAAGCA	
TCATTTCTGC	AAAAATGCAC	GTTAGCTTTC	TTCTTTCCTG	TAACTATTTG	8100
GTGAAGGGAA	GAGAAACTCT	CTAAGAGACT	GGCTCTGGAA	AATTGGTTGG	
GGGATTTTGA	GAACATCTTC	TTTTTTTTTT	TTTTTTTTTG	AGACAGAGTC	8200
TCACTCTGTT	GCCCAGGCTG	GAGTGTAGTG	GTGCAATCTT	GGCTCACTGC	
AACCTCCGCC	TCCCAGGTTT	AAGTGATTCT	CCTGCCTCAG	CCTCCTGAGT	8300
AGCTGGGATT	ACAGGTGTGC	ACCACCACGC	CAGGCTAATT	TTTTGTATTT	
TTAGTAGAGA	CGGGGGGGTC	TCACCAGTTT	GGCCAGCCTG	GTCTCGAACT	8400
CTGACTTCAG	GTGATCCACC	TGCCTCAGCC	TCCCAAAGTG	CTGGGATCAC	
AGGCGTGAGC	CACCGCGCCC	GGCGGGAACA	TCATTTTAAG	GGGATGTATC	8500
AGACATCTTT	ATGTTGCACT	TAGATTTAGG	AAATCTTTTG	GATACATTTT	
TATAAATGAG	AAGATTAAGT	TCTTATAGCT	CTCTAGTATC	TCAAAATCAT	8600
TGCCTGATTG	TTTGCAAAC	TGGTTTCTAG	CATGAAAGTC	TCAACTTCCC	
CATCAATGCC	ATTTGTCCTC	AGCTTCTCT	ATATGTTTCT	ACCACATCTG	8700
TGGTCATTTA	AAGTTGCCTA	CTGCTTGTGA	ACCCGGGAGG	TGGAGCTTGC	
AGTAAGCCGA	GATCGCGCCA	CTGCACTCCA	GCCTGAGCGA	CAGAGTGAGA	8800
CTCCATCTCA	AAAAAAAAAA	AAAAAAAAAG	TGCCTACTGC	CTTTGGTTTC	
CCAGATAACG	TGTCAAGTTT	CACCCTTGCC	CTCTTCAAAG	ATAACTGTAT	8900
TTTTTTTTTCC	TGGGTAGTTC	TCCGTATCAT	GCAAAAATAC	ATTGTATGTA	
GCTCCAAACT	GTACCTTTCA	TCTTCTAGT	CTTCTAAGA	GCATGGACCT	9000
AGTCTTTTTC	CTCTAAATAG	GGTAT			9025

FIGURE 1D

POLYMORPHISMS IN THE CODING SEQUENCE OF FY

ATGGCCTCCT	CTGGGTATGT	CCTCCAGGCG	GAGCTCTCCC	CCTCAACTGA	
GAACTCAAGT	CAGCTGGACT	TCGAAGATGT	ATGGAATTCT	TCCTATGGTG	100
TGAATGATTC	CTTCCCAGAT	GGAGACTATG	ATGCCAACCT	GGAAGCAGCT	
		G			
GCCCCCTGCC	ACTCCTGTAA	CCTGCTGGAT	GACTCTGCAC	TGCCCTTCTT	200
CATCCTCACC	AGTGTCCCTGG	GTATCCTAGC	TAGCAGCACT	GTCCTCTTCA	
T					
TGCTTTTCAG	ACCTCTCTTC	CGCTGGCAGC	TCTGCCCTGG	CTGGCCTGTC	300
		T			
CTGGCACAGC	TGGCTGTGGG	CAGTGCCCTC	TTCAGCATTG	TGGTGCCCGT	
A					
CTTGGCCCCA	GGGCTAGGTA	GCACTCGCAG	CTCTGCCCTG	TGTAGCCTGG	400
GCTACTGTGT	CTGGTATGGC	TCAGCCTTTG	CCCAGGCTTT	GCTGCTAGGG	
TGCCATGCCT	CCCTGGGGCA	CAGACTGGGT	GCAGGCCAGG	TCCCAGGCCT	500
CACCCTGGGG	CTCACTGTGG	GAATTTGGGG	AGTGGCTGCC	CTACTGACAC	
TGCCTGTCAC	CCTGGCCAGT	GGTGCTTCTG	GTGGACTCTG	CACCCTGATA	600
TACAGCACGG	AGCTGAAGGC	TTTGCAGGCC	ACACACACTG	TAGCCTGTCT	
TA					
TGCCATCTTT	GTCTTGTTGC	CATTGGGTTT	GTTTGGAGCC	AAGGGGCTGA	700
AGAAGGCATT	GGGTATGGGG	CCAGGCCCTT	GGATGAATAT	CCTGTGGGCC	
TGGTTTATTT	TCTGGTGGCC	TCATGGGGTG	GTTCTAGGAC	TGGATTTCTT	800
GGTGAGGTCC	AAGCTGTTGC	TGTTGTCAAC	ATGTCTGGCC	CAGCAGGCTC	
TGGACCTGCT	GCTGAACCTG	GCAGAAGCCC	TGGCAATTTT	GCACTGTGTG	900
GCTACGCCCC	TGCTCCTCGC	CCTATTCTGC	CACCAGGCCA	CCCGCACCCCT	
CTTGCCCTCT	CTGCCCCCTC	CTGAAGGATG	GTCTTCTCAT	CTGGACACCC	1000
		T			
TTGGAAGCAA	ATCCTAG				1017

FIGURE 2

ISOFORMS OF THE FY PROTEIN

MASSGYVLQA	ELSPSTENSS	QLDFEDVWNS	SYGVNDSFPD	GDYDANLEAA	
				G	
APCHSCNLLD	DSALPFFILT	SVLGILASST	VLFMLFRPLF	RWQLCPGWPV	100
	F			C	
LAQLAVGSAL	FSIVVPVLAP	GLGSTRSSAL	CSLGYCVWYG	SAFAQALLG	
T					
CHASLGHRLG	AGQVPGLTLG	LTVGIVGVAA	LLTLPVTLAS	GASGGLCTLI	200
YSELKALQA	THTVACLAIF	VLLPLGLFGA	KGLKKALGMG	PGPWMNILWA	
X					
WFIFWWPHGV	VLGLDFLVR	KLLLLSTCLA	QQALDLLNL	AEALAILHCV	300
ATPLLLALFC	HQATRTLLPS	LPLPEGWSSH	LDTLGSKS		338
		F			

X = M or I

FIGURE 3

FY_SEQ Listing TEMPLATE.ST25.txt
SEQUENCE LISTING

<110> Genaissance Pharmaceuticals, Inc.
Chew, Anne
Choi, Julie Y.
Koshy, Beena

<120> HAPLOTYPES OF THE FY GENE

<130> FY_MWH-1440PCT

<140> TBA

<141> 2001-10-12

<150> 60/240,275

<151> 2000-10-13

<160> 84

<170> PatentIn version 3.1

<210> 1

<211> 9025

<212> DNA

<213> Homo sapiens

<220>

<221> allele

<222> (2690)..(2690)

<223> PS1: polymorphic base C or T

<220>

<221> allele

<222> (2864)..(2864)

<223> PS2: polymorphic base G or A

<220>

<221> allele

<222> (2882)..(2882)

<223> PS3: polymorphic base A or G

<220>

<221> allele

<222> (2910)..(2910)

<223> PS4: polymorphic base C or T

<220>

<221> allele

<222> (2949)..(2949)

<223> PS5: polymorphic base C or A

<220>

<221> allele

<222> (2980)..(2980)

<223> PS6: polymorphic base G or C

<220>

<221> allele

FY_SEQ Listing TEMPLATE.ST25.txt

<222> (2996)..(2996)
<223> PS7: polymorphic base C or T

<220>
<221> allele
<222> (3259)..(3259)
<223> PS8: polymorphic base T or C

<220>
<221> allele
<222> (3470)..(3470)
<223> PS9: polymorphic base T or C

<220>
<221> allele
<222> (3672)..(3672)
<223> PS10: polymorphic base C or T

<220>
<221> allele
<222> (3707)..(3707)
<223> PS11: polymorphic base C or T

<220>
<221> allele
<222> (3979)..(3979)
<223> PS12: polymorphic base A or G

<220>
<221> allele
<222> (3997)..(3997)
<223> PS13: polymorphic base C or T

<220>
<221> allele
<222> (4140)..(4140)
<223> PS14: polymorphic base A or G

<220>
<221> allele
<222> (4214)..(4214)
<223> PS15: polymorphic base C or T

<220>
<221> allele
<222> (4280)..(4280)
<223> PS16: polymorphic base C or T

<220>
<221> allele
<222> (4313)..(4313)
<223> PS17: polymorphic base G or A

FY_SEQ Listing TEMPLATE.ST25.txt

<220>
 <221> allele
 <222> (4617)..(4617)
 <223> PS18: polymorphic base C or T

<220>
 <221> allele
 <222> (4618)..(4618)
 <223> PS19: polymorphic base G or A

<220>
 <221> allele
 <222> (4992)..(4992)
 <223> PS20: polymorphic base C or T

<400> 1
 ggcaaagggtt gggagtggtt tttcctcttg tagccacaca cctgagcact acggacaggg 60
 aggcaggtgc caccttgaca cctctcttcc atagcaatgg gaaagtgatg agtgcgggag 120
 tccctgaggag atgtggcctg cagacaacat gcagccatgc agggaccag gactgtaacc 180
 tggggaggag gcgggtccct gcaaggaaga gtagatttgg agaggaagga tggaggtgga 240
 ctctcaccoc attccccccg gaaatgaaca aagccggggc cttcccatag gaactgccct 300
 tggagatagc agagtgtggc tgcccctcct tgctccagca gcagtgggag aggcactgct 360
 ctggggcctg aactgcctct gcttcccccc ctgagggggc cctcactctt acccaagact 420
 ctggattgtt gcacggcaac cactcctccc atggcattgc tcagcaacta cttctccctt 480
 cccggccacc ctgtgcccc ttcttggtcc caacgccagc ccttcactct tccctccctca 540
 gcagccaggc agacataaca acaaaactac taaaaggagc ttcactgcag tgagctgttt 600
 cctgccccaa ctaagggaat aatgtgaact gtgtgcatgt gtgtggtgtg tatgcatgtg 660
 tgcatgtgtg tgtgtgtgtg tgcatgtgtg tgagtgtgtg agaggcagag cgagggaactg 720
 aggaggaggg ctaagagcca ggggtcctgg gcaagtggac agggctgtgg gacatgttgg 780
 ggaggctttg ggaatggggg attcctagtc aggggttcaca cctcacctgg gatgttgttc 840
 catgctggta tttcctctgc cacccccatt gcccatcggt cttggagaaa ggagtcctcc 900
 ggtgtgtgtt tgcccagctg tccattctat ctctccctta aacacagagc attcagccct 960
 tccctggatt tccctcctct gagccatgga gtcagtgcga cagcctttgc tatgcacctc 1020
 tcaggcctct ccttggcgtt gaccctggaa agacctaaca ccacctaatt tttcccatag 1080
 tctgtaccca gtgagttgaa ggctgggtcc ccacccctcc ttttgatttc ctgtcttctc 1140
 tctcgtggcc cagctgggtt gctgtggaga tgaggttcc ggtcctccct gtcctggctg 1200
 gactgccccg cctcagatcc aggatgccct tggcatogct cccacccctc cccagctttt 1260
 cctccctggt ctgacaatgg gcatgcaaaa aggggcagct gcaatctagc aggcctgccc 1320
 acccccttca gttcaggtaa tacagttgtg aatcttccag ccgctggtta gggccttggg 1380
 caccacagca agccccctac ctaagccggg gctactcct cttacaacag caagagagcc 1440
 ctggggcccc aggcctgttg agcttcttgt ctcccagcac ccgcttttgg gaaaatgact 1500
 tttcctcttc aagctgaacc actctgtcca tattacacag aagccatatt tgtacggggg 1560
 ggtgggaggg agaggggctg ttgtgctgtg tgtgtctgtc caggggtggg ggggtggggg 1620
 aagggagcag ggaggggacc gtgtatcttt ataactttc taactctcct gtgctaattc 1680
 cagaggggtc accctcaata tatctggatt atcctgttca ttcagctgcc tcttttctgg 1740
 tctcttggct gctgtgggga tgtgtgtatg tgagggtctt cttcccatag ccttgcacc 1800
 tgggtgcctg tgccataaaa ggtggtgtgt cccttgccag gccactctca agaatactta 1860
 tgtacagcaa caatataact ctacaaggga gagaagtgtg ttcacttctt tttgctaagc 1920
 ccttcccttc cagagagtgt cttggggggc atctgactgc ttccccccac cctctgccag 1980
 gcattgctgg agaattgtta gacggcgatg gagatgccat caacccccacc ctgcagagca 2040
 tcaccagaca ccaccagacc aaattcactt tccagccctt tcatgttgaa cctgaaactt 2100
 gagctagtgt cttgggagaa aagggggaaa tctctacgag gtacccatcc ttctgcacct 2160
 taggtctgag gtgcttggcc ccctaggaag ccctacatga atgggacaga aggtccttaa 2220
 caacactgga gatgaagcag ccgatgctgt tttggacaaa tgaaacagcg tcccctaacc 2280
 agccctttct atctcattgt tctgacttgg acacgccatg gctcaccgct cccaaagtcc 2340
 ccactagtct tccctagctg aggaataaaa ggtgatgaaa cagtgcagat 2400
 cctggggaaa cagctgagga ggggagggag ggggaagaag ccactaaaaa agtgaaatgt 2460
 gcttgggaga atcgccctgc ctgcagggtg gatgcccttt ctctctgctg gccagctctg 2520
 cccctcagtg agaaacttta catattgcta agatgcctgg ccaatgaaac agttccagag 2580
 actttatgtc ccccaagtag aatatgaata tgtgggcaat ggtcccattt 2640
 taaaaatgct tgtcccattg tcccctagag cctactttta cttgtcagay catgtattcc 2700

FY_SEQ Listing TEMPLATE.ST25.txt

acttcataatg	caagaggcat	gcactgagcc	cataggtggc	taggcaaaca	cccaatagct	2760
ccctgaaatg	gcttcattat	gcagcctcga	cagccacccc	aaccctccca	ctctcacact	2820
gaaacaccca	gacctagaga	tagctagaca	caccagaca	cccrccaagc	ccctcacata	2880
crगतatgtg	cacaatgata	cacagcaaa	gtacacagag	ttcagtacac	acaaagagct	2940
cacgcccang	tgcacacacc	cctcagttgg	gacagagttt	accaccacca	ccttttctcc	3000
aaacacatgg	cttttgaact	gcctttcctt	ggatccaggt	caaggggatg	gaggagcagt	3060
gagagtgcgc	cgcccttcca	ctccaatttc	ccagcacctc	ccttatctct	gcctcacaag	3120
tcacccagcc	ccctctctct	ccttccttgt	gcttgaagaa	tctctccttg	ctggaaagcc	3180
ccctgttttc	tcaactccc	tttccacttc	ggtaaaatct	ctacttgctg	gaaagccccc	3240
tgttttctca	atctccctyt	ccacttcggt	aaaatgcccc	ctttctgggt	cccacctttt	3300
tcttgagtgt	agtcccaacc	agccaaatcc	aacctcaaaa	caggaagacc	caaggccagt	3360
gacccccata	ggcctgaggg	ttgtgcaggg	agtgggcgtg	gggtaaggct	tcctgatgcc	3420
ccctgtccct	gcccagaacc	tgatggccct	cattagtcc	tggtcttay	cttggagaca	3480
caggcgctga	cagccgtccc	agccctctct	ccttgcggcc	tgaaccaaac	gggtgccatg	3540
ggaactgtct	gcacaggggt	agtatggggc	caggccccag	agtcccttat	ccctatgccc	3600
ctcatttccc	ctgctgtttg	cccctcagtc	tttatatctc	ttccttttcc	tcctcatctt	3660
ttctcccttc	cygctttttt	cctcttcctt	caaagtcttt	ttccttytct	ccttccatgt	3720
ctagccctct	agctccctct	tgtgtccctc	cctttcggct	tgagtgcagt	ccatccctgt	3780
ctcttggtgc	ctttccttct	gaccttgcac	tgtcctcca	gccccagctg	ccctggcttc	3840
cccaggactg	ttcctgtccc	ggctcttcag	gctccctgct	ttgtcctttt	ccactgtccc	3900
cactgcacct	gactcctgca	gagacctgt	tctcccaccc	gaccttctc	tctgtcctcc	3960
cctcccacct	gcccctcart	tcccaggaga	ccttctygg	gtaaactctga	tggcctctcc	4020
tggttatgtc	ctccaggcgg	agctctcccc	ctcaactgag	aactcaagtc	agctggactt	4080
cgaagatgta	tggaaattctt	cctatggtgt	gaatgattcc	ttcccagatg	gagactatgr	4140
tgccaaacctg	gaagcagctg	ccccctgcc	ctcctgtaac	ctgctggatg	actctgcact	4200
gccccttttc	atctgtacca	gtgtcctggg	tatcctagct	agcagcactg	tcctctctcat	4260
gcttttcaga	cctctcttcy	gctggcagct	ctgcccctgg	tggcctgtcc	tgrcacagct	4320
ggctgtgggc	agtgcctct	tcagcattgt	gggtcccgtc	ttggccccag	ggctaggtag	4380
cactgcagc	tctgcccctgt	gtagcctggg	ctactgtgtc	tggtatggct	cagcctttgc	4440
ccaggctttg	ctgctagggg	gccatgcctc	cctgggccac	agactgggtg	caggccctgt	4500
cccaggcctc	accctggggc	tcactgtggg	aatttgggga	gtggctgccc	tactgacact	4560
gcctgtcacc	ctggccagtg	gtgcttctgg	tggactctgc	accctgatat	acagcayrga	4620
gctgaaggct	ttgcaggcca	cacacactgt	agcctgtctt	gccatctttg	tcttgttgcc	4680
attgggtttg	ttcttgagcca	aggggctgaa	gaaggcattg	ggatgggggc	caggcccctg	4740
gatgaatatc	ctgtgggcct	ggtttatatt	ctgggtggcct	catgggggtg	ttctaggact	4800
ggatttcctg	gtgaggtcca	agctgttgct	gttgtcaaca	tgtctggccc	agcaggctct	4860
ggacctgtct	ctgaacctgg	cagaagccct	ggcaattttg	cactgtgtgg	ctacgcccct	4920
gctcctgcgc	ctgtctggcc	accaggccac	ccgcacccct	ttgcccctct	tgcccctccc	4980
tgaaggatgg	tyttctcatc	tggacaccct	tggaaagcaa	tcctagttct	cttcccacct	5040
gtcaacctga	attaaagtct	acactgcctt	tgtgaagcgg	gtgggtttct	attttgtctg	5100
gggagaagaa	ggagaatgga	gagagagaca	tttttatgtc	agactttctt	gccagtgtct	5160
gcttctatga	ctggcttggg	aagaagggtga	atgatgaata	aataccctca	gggtacacag	5220
atgttctctt	gaggtgtggg	gtcacggcca	tctcaaggga	gaagagaaga	ggaaccagag	5280
catgagggga	gtcattaaac	caaaaaaac	agaagggtat	gcttagctgg	aaaaaaagct	5340
gttctgggaa	gcaaatggaa	taggaactca	aactgagaga	taaacagtga	agagtgatga	5400
caaagcccag	agcaatacca	cctccccctg	tccaacctgc	ccagcctctg	tcttctgtct	5460
cctctctggc	tttgtttagt	gattaggaca	gtgggtggga	aggtgaaaga	agcatcccag	5520
gggatgttac	tcagttcagg	gaacatatca	aggtaattta	aaaagccact	tcctgggagt	5580
catctctccc	aggttcctca	gcctgacctg	aatgtgcgtg	cgtgtgtgtg	tgtgtgtgtg	5640
tgtacacatc	tgtttctcga	tctgcttagaa	tctaccttta	tgtagatgt	atgcatgtaa	5700
aaacatatgt	ccacccatga	gcttgcatct	ctgtcagcac	ctgaactgcg	cacacctgtg	5760
cgtgtgcact	gacttttctc	aggacccaaa	ccccactca	attctgcact	catccctgtt	5820
cacaggatat	agaatcgga	tttatgactc	actccttacc	caaatgagtt	ttctttaccc	5880
tggtttttaa	gcctagtctt	ttctgtgtag	gatgtgtgga	gggaagaaaa	gatcaagaag	5940
ttgtgaagg	tggagaaact	tgaaggggga	ggccctgatt	tgattcatct	tctgcttggg	6000
attccccgaa	tttccctttc	agaatctcag	cctttgaaat	aaacctttat	ttcccacata	6060
catctttcct	tcaccttccc	acacaatacc	ccaatcccc	gggcaccttt	ttcccacccc	6120
ctgattctct	ggctgcttaa	tcctgacctt	tgagattttt	ctcagtctct	acctacccaa	6180
gttttagatg	ctggaaggac	agaaacccct	cctcatcagg	ggcacagctt	ttaccaccaa	6240
gagcaaattc	accctctacc	caagaggcta	caaaacagtt	agttcctacc	tctaacccaa	6300
ctaaaggctg	gggaaacttg	agcagatacg	ttctatcagt	ttgaacccaa	ttaccatctt	6360
accattttcc	aaagatatgc	tatacctggt	ttctttacta	aaatgtttct	gcttgactct	6420
ctgggcttgg	gaatagtagg	cgagtgcggg	agagtgcag	agatgagtta	gaatagctta	6480
ggcaggagg	tgcaaaaggc	ttaggggaatt	ttcctgggtg	ggtgccacga	caaggcctct	6540

FY_SEQ Listing TEMPLATE.ST25.txt

aaatctccca	cctcctgtct	cttagcaacc	accaggttag	ctcctgattg	gttcgtctct	6600
aattgaaagg	cgggatttag	ggaccgattg	agacgcggga	gacattctga	aacagaaagg	6660
aagggagaga	aaatgaagag	aaaggaaata	atttacaaac	ctaaattatg	ctctggtttc	6720
caaccacagt	tcatgaatgt	gttctagtat	tttttcccc	gctttttttt	ttccaggctt	6780
ctctcaatat	ccccctccc	tccttgacca	ctcttgcaat	tctaccagat	gttgctgtcc	6840
tcctttacaa	ggtactgatt	tggaagctga	cctagttgag	ggggaggaga	gggcgttttt	6900
gactccctga	atcttccagt	gtcaacctga	tgcaagggag	gcttaattta	agaccagtag	6960
gcttgcttta	tctgccccca	accctgtgcc	tctggataga	aatccctggt	cagtcagtc	7020
agttagagag	aaccccgagc	tcctgggtta	tagcttgcca	gctctcatgg	ctttcacaag	7080
ggaaaggcag	ctgcagaagc	ccgaagctgc	taagaggtta	gggtgggctg	gagacagtc	7140
cctacccccg	ccccctgcta	catcctcctc	atccccaccc	ccaccgggat	tgctccaggc	7200
cttttgggct	gccctttccc	tgccattacc	taggcagcac	ttggagagct	cctccttaag	7260
tctaaccogg	acctcagtc	tttctttaaa	gctttcttgg	ggacctgcca	ccccatgcat	7320
ttaacccact	gcatgccatc	aaccactcta	aaattggtct	gagtcctggc	tcttttctgc	7380
aacccttcag	taatacaaat	cctgtctcct	aagaatttaa	tcttaggggt	7440	
ggcagggact	ttagctgtgt	atgagatatt	gggcatccta	gctaaagaaa	aaaatcctct	7500
cagaagatg	agagccagg	aagcaagctc	ttgggaaac	acaggaccct	gaggaaggtc	7560
agtttgcttt	gcttttctaa	ggagagagat	ctattattca	aggaagttt	gaacatcaca	7620
ttgacgctca	tagttcattt	attccaagct	gaggccctc	ccttaggatt	tagaaaacaa	7680
atacttggtc	ctcacaccct	ttttccattc	ctatttccct	atcccccaac	cccatcacca	7740
ccttctctcc	tcagaggaat	tctgattgag	aacttccact	ggatttcaaa	cccaattcat	7800
cgccaactct	aattgccaga	gatttgcatg	aaaaccatcg	tatgctatct	aattattctg	7860
acaaacagca	ccgcgcgtct	gggcacaagg	agaatcggag	ttttaattaa	caataatgca	7920
ccttgctgac	gaatgcgact	gttttaggtta	attaacaagt	ccaagtcctt	ccaaatcatc	7980
tctagacatc	taggtgattt	gggcaggaag	ggtgtgggga	acacagggag	ggatggggag	8040
tgtttaagca	tcattttctg	aaaaatgcac	gtagctttc	ttctttcctg	taactatttg	8100
gtgaaggaa	gagaaactct	ctaagagact	ggctctggaa	aattgggttg	gggattttga	8160
gaacatcttc	tttttttttt	tttttttttg	agacagagtc	tactctgtt	gcccaggctg	8220
gagtgtagtg	gtgcaatctt	ggctcactgc	aacctccgcc	tcccagggtt	aagtgattct	8280
cctgcctcag	cctcctgagt	agctgggatt	acagggtgtc	accaccacgc	caggctaatt	8340
ttttgtattt	ttagtagaga	cgggggggtc	tcaccagttt	ggccagcctg	gtctcgaaact	8400
ctgacttcag	gtgattccac	tgctcagcc	tcccaaagt	ctgggatcac	aggcgtgagc	8460
caccgcgccc	ggcggaacaa	tcattttaag	gggatgtatc	agacatcttt	atgttgcact	8520
tagatttagg	aaatcttttg	gatacatttt	tataaatgag	aagattaagt	tcttatagct	8580
ctctagatc	tcaaaactcat	tgctgtattg	tttgcaaaact	tggtttctag	catgaaaagtc	8640
tcaacttccc	catcaatgcc	atttgtcctc	agctttctct	atatgttctt	accacatctg	8700
tggtcattta	aagttgccta	ctgctgtgga	acccgggagg	tggagcttgc	agtaagccga	8760
gatcgcgcca	ctgcactcca	gcctgagcga	cagagtga	ctccatctca	aaaaaaaaaa	8820
aaaaaaaaagt	tgcaactgc	cttggttttc	ccagataaac	tgtaagttt	cacccttgcc	8880
ctcttcaag	ataactgtat	ttttttttcc	tgggtagttc	tccgtatcat	gcaaaaaatac	8940
attgtatgta	gtcctcaact	gtacctttca	tctttctagt	ctttctaaga	gcatggacct	9000
agtctttttc	ctctaaatag	ggtat				9025

<210> 2
 <211> 1017
 <212> DNA
 <213> Homo sapiens

<400> 2						
atggcctcct	ctgggtatgt	cctccaggcg	gagctctccc	cctcaactga	gaactcaagt	60
cagctgggact	tcgaagatgt	atggaattct	tcctatgggt	tgaatgattc	cttcccagat	120
ggagactatg	atgccaacct	ggaagcagct	gccccctgcc	actcctgtaa	cctgctggat	180
gactctgcac	tgcccttctt	catcctcacc	agtgtcctgg	gtatcctagc	tagcagcact	240
gtcctcttca	tgcttttcag	acctctcttc	cctggcagc	tctgcccctg	ctggcctgtc	300
ctggcacagc	tggctgtggg	cagtgccctc	ttcagcattg	tgggtcccgt	cttggcccca	360
gggctaggtta	gcaactgcag	ctctgccctg	tgtagcctgg	gctactgtgt	ctgggtatggc	420
tcagcctttg	cccaggcttt	gctgctaggg	tgccatgcct	ccctgggcca	cagactgggt	480
gcaggccagg	tcccaggcct	caccctgggg	ctcactgtgg	gaatttgggg	agtggtgcc	540
ctactgacac	tgccctgtcac	cctggccagt	gtgcttcttg	gtggactctg	caccctgata	600
tacagcacgg	agctgaaggc	tttgaggcc	acacacactg	tagcctgtct	tgccatcttt	660
gtcttggtgc	cattgggttt	gtttggagcc	aaggggctga	agaaggcatt	gggtatgggg	720
ccaggccctt	ggatgaatat	cctgtgggcc	tgggttattt	tctggtggcc	tcattgggtg	780
gttctaggac	tggatttcct	ggtgaggtcc	aagctgttgc	tgttgtcaac	atgtctggcc	840

FY_SEQ Listing TEMPLATE.ST25.txt

```

cagcaggctc tggacctgct gctgaacctg gcagaagccc tggcaatttt gcaactgtgtg 900
gctacgcccc tgctcctcgc cctattctgc caccaggcca cccgcaccct cttgccctct 960
ctgcccctcc ctgaaggatg gtcttctcat ctggacaccc ttggaagcaa atcctag 1017

```

```

<210> 3
<211> 338
<212> PRT
<213> Homo sapiens

```

```

<400> 3

```

```

Met Ala Ser Ser Gly Tyr Val Leu Gln Ala Glu Leu Ser Pro Ser Thr
1      5      10      15
Glu Asn Ser Ser Gln Leu Asp Phe Glu Asp Val Trp Asn Ser Ser Tyr
20      25      30
Gly Val Asn Asp Ser Phe Pro Asp Gly Asp Tyr Asp Ala Asn Leu Glu
35      40      45
Ala Ala Ala Pro Cys His Ser Cys Asn Leu Leu Asp Asp Ser Ala Leu
50      55      60
Pro Phe Phe Ile Leu Thr Ser Val Leu Gly Ile Leu Ala Ser Ser Thr
65      70      75      80
Val Leu Phe Met Leu Phe Arg Pro Leu Phe Arg Trp Gln Leu Cys Pro
85      90      95
Gly Trp Pro Val Leu Ala Gln Leu Ala Val Gly Ser Ala Leu Phe Ser
100     105     110
Ile Val Val Pro Val Leu Ala Pro Gly Leu Gly Ser Thr Arg Ser Ser
115     120     125
Ala Leu Cys Ser Leu Gly Tyr Cys Val Trp Tyr Gly Ser Ala Phe Ala
130     135     140
Gln Ala Leu Leu Leu Gly Cys His Ala Ser Leu Gly His Arg Leu Gly
145     150     155     160
Ala Gly Gln Val Pro Gly Leu Thr Leu Gly Leu Thr Val Gly Ile Trp
165     170     175
Gly Val Ala Ala Leu Leu Thr Leu Pro Val Thr Leu Ala Ser Gly Ala
180     185     190
Ser Gly Gly Leu Cys Thr Leu Ile Tyr Ser Thr Glu Leu Lys Ala Leu
195     200     205
Gln Ala Thr His Thr Val Ala Cys Leu Ala Ile Phe Val Leu Leu Pro
210     215     220
Leu Gly Leu Phe Gly Ala Lys Gly Leu Lys Lys Ala Leu Gly Met Gly
225     230     235     240
Pro Gly Pro Trp Met Asn Ile Leu Trp Ala Trp Phe Ile Phe Trp Trp
245     250     255
Pro His Gly Val Val Leu Gly Leu Asp Phe Leu Val Arg Ser Lys Leu
260     265     270
Leu Leu Leu Ser Thr Cys Leu Ala Gln Gln Ala Leu Asp Leu Leu Leu

```

FY_SEQ Listing TEMPLATE.ST25.txt
280 285

275

Asn Leu Ala Glu Ala Leu Ala Ile Leu His Cys Val Ala Thr Pro Leu
 290 295 300

Leu Leu Ala Leu Phe Cys His Gln Ala Thr Arg Thr Leu Leu Pro Ser
 305 310 315 320

Leu Pro Leu Pro Glu Gly Trp Ser Ser His Leu Asp Thr Leu Gly Ser
 325 330 335

Lys Ser

<210> 4
 <211> 15
 <212> DNA
 <213> Homo sapiens

<400> 4
 tgtcagayca tgtat 15

<210> 5
 <211> 15
 <212> DNA
 <213> Homo sapiens

<400> 5
 gacaccrcc aagcc 15

<210> 6
 <211> 15
 <212> DNA
 <213> Homo sapiens

<400> 6
 cacatacrga tatgt 15

<210> 7
 <211> 15
 <212> DNA
 <213> Homo sapiens

<400> 7
 cagcaaaygt acaca 15

<210> 8
 <211> 15
 <212> DNA
 <213> Homo sapiens

<400> 8
 acgcccangt gcaca 15

<210> 9
 <211> 15
 <212> DNA
 <213> Homo sapiens

FY_SEQ Listing TEMPLATE.ST25.txt

<400> 9
cagagttsac cacca 15

<210> 10
<211> 15
<212> DNA
<213> Homo sapiens

<400> 10
caccttityc ccaaa 15

<210> 11
<211> 15
<212> DNA
<213> Homo sapiens

<400> 11
tctccctytc cactt 15

<210> 12
<211> 15
<212> DNA
<213> Homo sapiens

<400> 12
cccttcgygc ttttt 15

<210> 13
<211> 15
<212> DNA
<213> Homo sapiens

<400> 13
tttccttytc tcctt 15

<210> 14
<211> 15
<212> DNA
<213> Homo sapiens

<400> 14
cccctcartt cccag 15

<210> 15
<211> 15
<212> DNA
<213> Homo sapiens

<400> 15
actcttcygg tgtaa 15

<210> 16
<211> 15
<212> DNA
<213> Homo sapiens

FY_SEQ Listing TEMPLATE.ST25.txt

<400> 16
cttcatcytc accag 15

<210> 17
<211> 15
<212> DNA
<213> Homo sapiens

<400> 17
tacagcaygg agctg 15

<210> 18
<211> 15
<212> DNA
<213> Homo sapiens

<400> 18
acagcacrga gctga 15

<210> 19
<211> 15
<212> DNA
<213> Homo sapiens

<400> 19
ggatggytt ctcac 15

<210> 20
<211> 15
<212> DNA
<213> Homo sapiens

<400> 20
ttaacttgac agayc 15

<210> 21
<211> 15
<212> DNA
<213> Homo sapiens

<400> 21
agtgaatac atgrt 15

<210> 22
<211> 15
<212> DNA
<213> Homo sapiens

<400> 22
caccagaca cccac 15

<210> 23
<211> 15
<212> DNA
<213> Homo sapiens

<400> 23

FY_SEQ Listing TEMPLATE.ST25.txt

gtgaggggct tgggyg 15

<210> 24
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 24
gccctcacac tacrg 15

<210> 25
<211> 15
<212> DNA
<213> Homo sapiens

<400> 25
ttgtgcacat atcyg 15

<210> 26
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 26
gatacacagc aaayg 15

<210> 27
<211> 15
<212> DNA
<213> Homo sapiens

<400> 27
gaactctgtg tacrt 15

<210> 28
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 28
gagctcacgc ccamg 15

<210> 29
<211> 15
<212> DNA
<213> Homo sapiens

<400> 29
ggggtgtgtg cackt 15

<210> 30
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 30
ttgggacaga gttsa 15

FY_SEQ Listing TEMPLATE.ST25.txt

<210> 31
<211> 15
<212> DNA
<213> Homo sapiens

<400> 31
aggtggtggt ggtsa 15

<210> 32
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 32
caccaccacc tttyt 15

<210> 33
<211> 15
<212> DNA
<213> Homo sapiens

<400> 33
catgtgtttg ggara 15

<210> 34
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 34
tctcaatctc cctyt 15

<210> 35
<211> 15
<212> DNA
<213> Homo sapiens

<400> 35
ttaccgaagt ggara 15

<210> 36
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 36
ttttctccct tccyg 15

<210> 37
<211> 15
<212> DNA
<213> Homo sapiens

<400> 37
agaggaaaaa agcrg 15

FY_SEQ Listing TEMPLATE.ST25.txt

<210> 38
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 38
agtcttttttc cttyt 15

<210> 39
<211> 15
<212> DNA
<213> Homo sapiens

<400> 39
cataggaagg agara 15

<210> 40
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 40
cacctgcccc tcart 15

<210> 41
<211> 15
<212> DNA
<213> Homo sapiens

<400> 41
agtctcctgg gaayt 15

<210> 42
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 42
caggagactc ttcyg 15

<210> 43
<211> 15
<212> DNA
<213> Homo sapiens

<400> 43
tcagagttac accrg 15

<210> 44
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 44
gcccttcttc atcyt 15

FY_SEQ Listing TEMPLATE.ST25.txt

<210> 45
<211> 15
<212> DNA
<213> Homo sapiens

<400> 45
aggacactgg tgarg

15

<210> 46
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 46
ctgatataca gcayg

15

<210> 47
<211> 15
<212> DNA
<213> Homo sapiens

<400> 47
agccttcagc tcrt

15

<210> 48
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 48
tgatatacag cacrg

15

<210> 49
<211> 15
<212> DNA
<213> Homo sapiens

<400> 49
aagccttcag ctcyg

15

<210> 50
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 50
cctgaaggat ggtyt

15

<210> 51
<211> 15
<212> DNA
<213> Homo sapiens

<400> 51
gtccagatga gaara

15

<210> 52

FY_SEQ Listing TEMPLATE.ST25.txt

<211> 10
<212> DNA
<213> Homo sapiens

<400> 52
acttgtcaga 10

<210> 53
<211> 10
<212> DNA
<213> Homo sapiens

<400> 53
ggaatacatg 10

<210> 54
<211> 10
<212> DNA
<213> Homo sapiens

<400> 54
ccagacaccc 10

<210> 55
<211> 10
<212> DNA
<213> Homo sapiens

<400> 55
aggggcttgg 10

<210> 56
<211> 10
<212> DNA
<213> Homo sapiens

<400> 56
cctcacatac 10

<210> 57
<211> 10
<212> DNA
<213> Homo sapiens

<400> 57
tgcacatatc 10

<210> 58
<211> 10
<212> DNA
<213> Homo sapiens

<400> 58
acacagcaaa 10

<210> 59
<211> 10

FY_SEQ Listing TEMPLATE.ST25.txt

<212> DNA
<213> Homo sapiens

<400> 59
ctctgtgtac 10

<210> 60
<211> 10
<212> DNA
<213> Homo sapiens

<400> 60
ctcacgccca 10

<210> 61
<211> 10
<212> DNA
<213> Homo sapiens

<400> 61
gtgtgtgcac 10

<210> 62
<211> 10
<212> DNA
<213> Homo sapiens

<400> 62
ggacagagtt 10

<210> 63
<211> 10
<212> DNA
<213> Homo sapiens

<400> 63
tggtggtggt 10

<210> 64
<211> 10
<212> DNA
<213> Homo sapiens

<400> 64
caccaccttt 10

<210> 65
<211> 10
<212> DNA
<213> Homo sapiens

<400> 65
gtgtttggga 10

<210> 66
<211> 10
<212> DNA

FY_SEQ-Listing TEMPLATE.ST25.txt

<213> Homo sapiens

<400> 66
caatctccct

10

<210> 67
<211> 10
<212> DNA
<213> Homo sapiens<400> 67
ccgaagtgga

10

<210> 68
<211> 10
<212> DNA
<213> Homo sapiens<400> 68
tctcccttcc

10

<210> 69
<211> 10
<212> DNA
<213> Homo sapiens<400> 69
ggaaaaaagc

10

<210> 70
<211> 10
<212> DNA
<213> Homo sapiens<400> 70
ctttttcctt

10

<210> 71
<211> 10
<212> DNA
<213> Homo sapiens<400> 71
aggaaggaga

10

<210> 72
<211> 10
<212> DNA
<213> Homo sapiens<400> 72
ctgcccctca

10

<210> 73
<211> 10
<212> DNA
<213> Homo sapiens

FY_SEQ Listing TEMPLATE.ST25.txt

<400> 73
ctcctgggaa 10

<210> 74
<211> 10
<212> DNA
<213> Homo sapiens

<400> 74
gagactcttc 10

<210> 75
<211> 10
<212> DNA
<213> Homo sapiens

<400> 75
gagttacacc 10

<210> 76
<211> 10
<212> DNA
<213> Homo sapiens

<400> 76
cttcttcac 10

<210> 77
<211> 10
<212> DNA
<213> Homo sapiens

<400> 77
acactggtga 10

<210> 78
<211> 10
<212> DNA
<213> Homo sapiens

<400> 78
atatacagca 10

<210> 79
<211> 10
<212> DNA
<213> Homo sapiens

<400> 79
cttcagctcc 10

<210> 80
<211> 10
<212> DNA
<213> Homo sapiens

FY_SEQ Listing TEMPLATE.ST25.txt

<400> 80
tatacagcac 10

<210> 81
<211> 10
<212> DNA
<213> Homo sapiens

<400> 81
ccttcagctc 10

<210> 82
<211> 10
<212> DNA
<213> Homo sapiens

<400> 82
gaaggatggt 10

<210> 83
<211> 10
<212> DNA
<213> Homo sapiens

<400> 83
cagatgagaa 10

<210> 84
<211> 2400
<212> DNA
<213> Homo sapiens

<220>
<221> allele
<222> (30)..(30)
<223> PS1: polymorphic base C or T

<220>
<221> misc_feature
<222> (61)..(120)
<223> Ns represent sequence between PS

<220>
<221> allele
<222> (150)..(150)
<223> PS2: polymorphic base G or A

<220>
<221> misc_feature
<222> (181)..(240)
<223> Ns represent sequence between PS

<220>
<221> allele
<222> (270)..(270)
<223> PS3: polymorphic base A or G

FY_SEQ Listing TEMPLATE.ST25.txt

<220>
<221> misc_feature
<222> (301)..(360)
<223> Ns represent sequence between PS

<220>
<221> allele
<222> (390)..(390)
<223> PS4: polymorphic base C or T

<220>
<221> misc_feature
<222> (421)..(480)
<223> Ns represent sequence between PS

<220>
<221> allele
<222> (510)..(510)
<223> PS5: polymorphic base C or A

<220>
<221> misc_feature
<222> (541)..(600)
<223> Ns represent sequence between PS

<220>
<221> allele
<222> (630)..(630)
<223> PS6: polymorphic base G or C

<220>
<221> misc_feature
<222> (661)..(720)
<223> Ns represent sequence between PS

<220>
<221> allele
<222> (750)..(750)
<223> PS7: polymorphic base C or T

<220>
<221> misc_feature
<222> (781)..(840)
<223> Ns represent sequence between PS

<220>
<221> allele
<222> (870)..(870)
<223> PS8: polymorphic base T or C

<220>
<221> misc_feature

FY_SEQ Listing TEMPLATE.ST25.txt

<222> (901)..(960)
<223> Ns represent sequence between PS

<220>
<221> allele
<222> (990)..(990)
<223> PS9: polymorphic base T or C

<220>
<221> misc_feature
<222> (1021)..(1080)
<223> Ns represent sequence between PS

<220>
<221> allele
<222> (1110)..(1110)
<223> PS10: polymorphic base C or T

<220>
<221> misc_feature
<222> (1141)..(1200)
<223> Ns represent sequence between PS

<220>
<221> allele
<222> (1230)..(1230)
<223> PS11: polymorphic base C or T

<220>
<221> misc_feature
<222> (1261)..(1320)
<223> Ns represent sequence between PS

<220>
<221> allele
<222> (1350)..(1350)
<223> PS12: polymorphic base A or G

<220>
<221> misc_feature
<222> (1381)..(1440)
<223> Ns represent sequence between PS

<220>
<221> allele
<222> (1470)..(1470)
<223> PS13: polymorphic base C or T

<220>
<221> misc_feature
<222> (1501)..(1560)
<223> Ns represent sequence between PS

FY_SEQ Listing TEMPLATE.ST25.txt

<220>
<221> allele
<222> (1590)..(1590)
<223> PS14: polymorphic base A or G

<220>
<221> misc_feature
<222> (1621)..(1680)
<223> Ns represent sequence between PS

<220>
<221> allele
<222> (1710)..(1710)
<223> PS15: polymorphic base C or T

<220>
<221> misc_feature
<222> (1741)..(1800)
<223> Ns represent sequence between PS

<220>
<221> allele
<222> (1830)..(1830)
<223> PS16: polymorphic base C or T

<220>
<221> misc_feature
<222> (1861)..(1920)
<223> Ns represent sequence between PS

<220>
<221> allele
<222> (1950)..(1950)
<223> PS17: polymorphic base G or A

<220>
<221> misc_feature
<222> (1981)..(2040)
<223> Ns represent sequence between PS

<220>
<221> allele
<222> (2070)..(2070)
<223> PS18: polymorphic base G or A

<220>
<221> misc_feature
<222> (2101)..(2160)
<223> Ns represent sequence between PS

<220>
<221> allele
<222> (2190)..(2190)
<223> PS19: polymorphic base G or A

FY_SEQ Listing TEMPLATE.ST25.txt

<220>
 <221> misc_feature
 <222> (2221)..(2280)
 <223> Ns represent sequence between PS

<220>
 <221> allele
 <222> (2310)..(2310)
 <223> PS20: polymorphic base C or T

<220>
 <221> misc_feature
 <222> (2341)..(2400)
 <223> Ns represent sequence 3' to PS20

<400> 84
 tcccctagag cctacttta cttgtcagay catgtattcc acttcatatg caagaggcat 60
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 120
 tagagatagc tagacacacc cagacacccr ccaagcccct caçatacaga tatgtgcaca 180
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240
 cccagacacc cgccaagccc ctacatacr gatagtgc caatgataca cagcaaacgt 300
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360
 cagatatgtg cacaatgata cacagcaaay gtacacagag ttcagtacac acaaagagct 420
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480
 gttcagtaca cacaaagagc tcacgcccac gtgcacacac ccctcagttg ggacagagtt 540
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 600
 tgcacacacc cctcagttgg gacagagttt accaccacca cctttctccc aaacacatgg 660
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 720
 ttgggacaga gttgaccacc accaccttty tcccaaacac atggcctttt aactgccttt 780
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 840
 ggaaagcccc ctgttttctc aatctcccty tccacttcgg taaaatgcc actttctggt 900
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 960
 tgatggcctc ttgctcttct tggctcttay ctgggaagca caggcgctga cagcgctccc 1020
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1080
 ccttttcttc ctcatctttt ctcccttccty gcttttttcc tcttctctca aagtcttttt 1140
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1200
 ttctctcttc cttcaaagtc tttttctcty tctccttctc atgctagcct cctagctccc 1260
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1320
 ctctgtcttc cctcccacc tgccctcar ttcccaggag actcttccgg tgtaactctg 1380
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1440
 cctgcccctc aattcccagg agactcttct ggtgtaactc tgatggcctc ctctgggtat 1500
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1560
 gaatgattcc ttcccagatg gagactatgr tgccaacctg gaagcagctg ccccctgcca 1620
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1680
 tggatgactc tgactgccc ttcttcacat tcaccagtgt cctgggtatc ctactagca 1740
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1800
 tctcttctcat gcttttcaga cctctcttcty gctggcagct ctgccctggc tggcctgtcc 1860
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1920
 ggcagctctg cctggctgg cctgtctcgr cacagctggc tgtgggcagt gccctcttca 1980
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2040
 tgggtggactc tgcaccctga tatacagcay ggagctgaag gctttgcagg ccacacacac 2100
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2160
 ggtggactct gcaccctgat atacagcacr gagctgaagg ctttgcaggc cacacacact 2220
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2280
 gccctctctg ccctccctg aaggatggtt ttctcatctg gacacccttg gaagcaaadc 2340
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2400

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.